

BAYESIAN SEMI-BLIND COMPONENT SEPARATION FOR FOREGROUND REMOVAL IN
INTERFEROMETRIC 21-CM OBSERVATIONSLE ZHANG¹, EMORY F. BUNN⁷, ATA KARAKCI², ANDREI KOROTKOV², P.M. SUTTER^{9,10,6}, PETER T. TIMBIE¹,
GREGORY S. TUCKER², BENJAMIN D. WANDELT^{4,5,3,8}¹DEPARTMENT OF PHYSICS, UNIVERSITY OF WISCONSIN, MADISON, WI 53706, USA²DEPARTMENT OF PHYSICS, BROWN UNIVERSITY, 182 HOPE STREET, PROVIDENCE, RI 02912, USA³DEPARTMENT OF PHYSICS, 1110 W GREEN STREET, UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN, URBANA, IL 61801, USA⁴UPMC UNIV PARIS 06, UMR 7095, INSTITUT D'ASTROPHYSIQUE DE PARIS, 98 BIS, BOULEVARD ARAGO, 75014 PHARIS, FRANCE⁵CNRS, UMR 7095, INSTITUT D'ASTROPHYSIQUE DE PARIS, 98 BIS, BOULEVARD ARAGO, 75014 PARIS, FRANCE⁶CENTER FOR COSMOLOGY AND ASTRO-PARTICLE PHYSICS, OHIO STATE UNIVERSITY, COLUMBUS, OH 43210, USA⁷PHYSICS DEPARTMENT, UNIVERSITY OF RICHMOND, RICHMOND, VIRGINIA 23173, USA⁸DEPARTMENT OF ASTRONOMY, UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN, URBANA, IL 61801, USA⁹INFN - NATIONAL INSTITUTE FOR NUCLEAR PHYSICS, VIA VALERIO 2, I-34127 TRIESTE, ITALY¹⁰INAF - OSSERVATORIO ASTRONOMIC DI TRIESTE, VIA TIEPOLO 11, I-34143 TRIESTE, ITALY*Draft version January 12, 2016*

ABSTRACT

We present in this paper a new Bayesian semi-blind approach for foreground removal in observations of the 21-cm signal with interferometers. The technique, which we call HIEMICA (HI Expectation-Maximization Independent Component Analysis), is an extension of the Independent Component Analysis (ICA) technique developed for two-dimensional (2D) CMB maps to three-dimensional (3D) 21-cm cosmological signals measured by interferometers. This technique provides a fully Bayesian inference of power spectra and maps and separates the foregrounds from signal based on the diversity of their power spectra. Only relying on the statistical independence of the components, this approach can jointly estimate the 3D power spectrum of the 21-cm signal and, the 2D angular power spectrum and the frequency dependence of each foreground component, without any prior assumptions about foregrounds. This approach has been tested extensively by applying it to mock data from interferometric 21-cm intensity mapping observations under idealized assumptions of instrumental effects. We also discuss the impact when the noise properties are not known completely. As a first step toward solving the 21 cm power spectrum analysis problem we compare the semi-blind HIEMICA technique with the commonly used Principal Component Analysis (PCA). Under the same idealized circumstances the proposed technique provides significantly improved recovery of the power spectrum. This technique can be applied straightforwardly to all 21-cm interferometric observations, including epoch of reionization measurements, and can be extended to single-dish observations as well.

Subject headings: cosmology:observations, 21-cm, foreground, instrumentation:interferometers, methods: data analysis, statistical

1. INTRODUCTION

One of the main goals of modern cosmology is to measure the large-scale structure (LSS) of the universe, which encodes crucial information about cosmological processes. Measuring these large-scale fluctuations is the primary tool for understanding cosmological and astrophysical phenomena and determining related parameters. Observations (Planck Collaboration et al. 2014b; Hinshaw et al. 2013; Smoot et al. 1992) of the cosmic microwave background (CMB) precisely probe fluctuations induced by primordial matter density perturbations from the last scattering surface at a redshift of $z \sim 1100$ to constrain models of inflation (Peiris et al. 2003; Kinney et al. 2006). Spectroscopic galaxy redshift surveys such as the 2dF Galaxy Redshift Survey (Colless et al. 2003), WiggleZ Dark Energy Survey (Drinkwater et al. 2010), the Sloan Digital Sky Survey (York et al. 2000), the LAMOST (Wang et al. 2009) and Baryon Oscillation

Spectroscopic Survey (Anderson et al. 2013) have been attempting to map out the three-dimensional (3D) structure of galaxies in the universe, providing an even more precise measurement of cosmological parameters.

Neutral hydrogen (HI) tomography may provide a powerful alternative to galaxy surveys. The redshifted 21-cm hyperfine transition line allows a direct measurement of the redshift (distance) of a source. Using HI tomography through the redshifted 21-cm emissions could be a relatively inexpensive tool to probe LSS over enormous volumes of our universe, from redshift 0 to ultimately as high as ~ 200 , potentially collecting orders of magnitude more information than either the CMB or galaxy surveys can provide. Cosmological parameters could be determined with far greater precision (Mao et al. 2008), and the properties of dark matter, dark energy, neutrinos (Pritchard & Loeb 2012; Morales & Wyithe 2010; Loeb & Wyithe 2008), and non-Gaussianity (Maldacena 2003; Loeb & Zaldarriaga 2004; Xu et al. 2015) from inflation could be tightly constrained. 21-cm emis-

sion may be the only way to explore the dark ages and the epoch of reionization (Madau et al. 1997; Chen & Miralda-Escude 2004; McQuinn et al. 2006; Furlanetto et al. 2006).

In particular, it is possible to trace the dynamics of the expansion of the universe and study the evolution of the dark energy by mapping the large scale structure of the universe in three dimensions (Seo & Eisenstein 2003). Several authors (Morales & Wyithe 2010; Chang et al. 2008; Wyithe et al. 2008; Loeb & Wyithe 2008; Peterson et al. 2006) have proposed a new technique, the so-called “21-cm intensity mapping”, for the observation of the baryon acoustic oscillations (BAO) in the large scale structure. By mapping the 21-cm intensity field with an angular resolution of ~ 10 arcmin across the redshift range from 0 up to 3, it may be possible to determine precisely the equation of state of the dark energy (Bull et al. 2015). The technique has been tested with existing telescopes such as the Green Bank Telescope (GBT) (Chang et al. 2010; Masui et al. 2013; Switzer et al. 2013) and the Parkes Radio Telescope (Pen et al. 2009). However, to obtain the sensitivity and angular resolution necessary to resolve the BAO peaks demands dedicated radio telescopes with large numbers of receivers and apertures of order 100 m. Dedicated single-aperture telescopes (e.g. FAST (Nan et al. 2011)) are prohibitively expensive. Instead, interferometers provide an economical way to observe large volumes of the sky and measure the BAO features efficiently. Additionally, interferometry is well-suited to the EoR and BAO measurements: visibilities naturally measure Fourier modes of the intensity distribution on sky, which is directly related to the power spectrum of statistical fluctuations in the HI distribution. Radio interferometric arrays are already being designed and constructed. Currently, there are several dedicated HI interferometers that plan to detect the redshifted 21-cm EoR signal: LOFAR¹, MWA², 21-cmA³ and PAPER⁴. GMRT⁵ is also currently under way to detect the 21 cm signal from reionization. Furthermore, several 21-cm BAO experiments have been proposed (including BAOBAB (Pober et al. 2013a), BAO-Radio (Ansari et al. 2012b)) or constructed (CHIME⁶, and TianLai (Chen 2012)) to conduct low redshift surveys. Next-generation experiments such as SKA⁷ will enable revolutionary progress in 21-cm tomography.

However, HI tomography is extremely challenging because astrophysical foreground contamination is expected to be at least four orders of magnitude brighter than the 21-cm signal. The success of statistical detection of the signal relies on robust foreground removal, which has been extensively explored over the past decade. Different methods have different advantages and biases in providing signal estimates. Fortunately, foregrounds are expected to have smooth continuum spectra that vary slowly with frequency, while the 21-cm signal is expected to oscillate dramatically over a short frequency range and become uncorrelated beyond frequency separa-

tions of about 1 MHz (Santos et al. 2005). In principle, one can remove slowly varying components to extract the HI signal from the contaminants although this process will always leave some residual contamination. The early ideas use a multi-frequency angular cross-correlation power spectrum to separate foregrounds from the 21-cm signal (Di Matteo et al. 2002, 2004; Oh & Mack 2003; Zaldarriaga et al. 2004; Santos et al. 2005; Cooray & Furlanetto 2004). An alternative approach is to subtract foregrounds along each line of sight in the sky or in the uv -plane, by fitting the frequency dependence with a particular smooth function (e.g. a polynomial), and then using the residuals to determine the 21-cm power spectrum (Santos et al. 2005; Wang et al. 2006; McQuinn et al. 2006; Bowman et al. 2006; Jelić et al. 2008; Gleser et al. 2008; Liu et al. 2009a; Harker et al. 2009a; Petrovic & Oh 2011; Bowman et al. 2009; Harker et al. 2010). Higher-order statistics, like skewness, are also employed to clean foregrounds (Harker et al. 2009b). In addition, Liu et al. (2009b); Trott et al. (2012); Morales et al. (2006) examined the impact of point-source subtraction residuals on 21-cm fluctuations, and Morales & Hewitt (2004) found a removal approach based on some symmetries in the 3D Fourier representation of radio interferometric data. More recently, simulations and observations have shown that even if sky foregrounds are spectrally smooth, the chromatic synthesized beam of an interferometer interacting with smooth spectrum foregrounds can still imprint unsmooth spectral features on measured foregrounds, contaminating a “wedge”-like region (so-called “mode-mixing”) in cylindrical $(k_{\perp}, k_{\parallel})$ Fourier space (Datta et al. 2010; Vedantham et al. 2012; Morales et al. 2012; Parsons et al. 2012; Trott et al. 2012; Pober et al. 2013a,b; Dillon et al. 2013; Liu et al. 2014a,b; Pober et al. 2014).

In order to subtract foregrounds with minimal prior assumptions on the angular distribution or frequency dependence of foreground components, some *blind* (non-parametric) methods have been extensively studied in the literature. By using only the data, those methods can automatically find the correlated features and separate them out into distinct components. A commonly used data analysis technique known as Principal Component Analysis (PCA) has been applied to foreground removal and successfully tested on simulated data and real observations (de Oliveira-Costa et al. 2008; Chang et al. 2010; Masui et al. 2013; Switzer et al. 2013). Other non-parametric methods, including the FASTICA algorithm (Hyvärinen 1999; Hyvärinen & Oja 2000; Chapman et al. 2012; Wolz et al. 2014) based on independent component analysis and singular value decomposition (SVD) together with an analysis based on the Karhunen-Loève transform (Shaw et al. 2014), provide encouraging results. The FASTICA algorithm is based on the maximization of non-Gaussianity (neg-entropy), a measure of the deviation of a mixture of signals from a Gaussian distribution. However, the disadvantage of this approach is that the noise leakage into the reconstructed signal is in principle hard to estimate since the signal and noise essentially have the same statistical distribution, which is completely uncorrelated across frequencies. Other methods such as the Correlated Component Analysis (CCA) (Bonaldi & Brown 2015) can be referred to as “semi-blind” as they require some prior knowledge. By

¹ <http://www.lofar.org>

² <http://www.mwatelescope.org>

³ <http://21-cma.bao.ac.cn>

⁴ <http://astro.berkeley.edu/dbacker/eor/>

⁵ <http://gmrt.ncra.tifr.res.in>

⁶ <http://chime.phas.ubc.ca>

⁷ <http://www.skatelescope.org>

using second-order statistics CCA also provides a promising way to separate the 21-cm reionization signal from foreground contamination for SKA measurements. Recently Vansyngel et al. (2014) have provided a unified approach for semi-blind foreground cleaning from multi-frequency CMB data.

In this study, we describe HIEMICA, the HI Expectation-Maximization Independent Component Analysis algorithm, a Bayesian framework to infer power spectra and maps of the HI signal and foregrounds from “dirty” data cubes. This technique extends the Spectral Matching ICA (SMICA) (Cardoso et al. 2008; Delabrouille et al. 2003; Planck Collaboration et al. 2014a) approach from two-dimensional (2D) CMB maps to the three-dimensional (3D) 21-cm cosmological signal and can be considered a semi-blind source separation method. The statistical isotropy of the cosmological Gaussian random fields in 3D Fourier space provides a characteristic and unique signature to distinguish them from the statistically non-isotropic foregrounds; in the frequency (redshift, or radial) direction, astrophysical foreground sources contain only spectral information, which has no connection with their spatial distributions in the transverse direction.

The purpose of this work is to take a first step toward building a new framework for 21cm data analysis. Our paper provides a complete description of the mathematical formalism and numerical techniques, which in principle can be applied to any realistic array design. Here we test the approach for an idealized array; we propose to dedicate a future paper to systematic tests of arrays with realistic instrumental effects.

This paper is organized as follows: Sect. 2 describes the HIEMICA semi-blind component separation algorithm; Sect. 3 summarizes details of our performed simulations; Sect. 4 presents the application of HIEMICA to simulated sky maps and shows the main results. Finally, Sect. 5 provides discussion and concluding remarks.

2. THE HIEMICA METHOD

2.1. Radio interferometric measurements

For a multichannel interferometric array, the fundamental observable is the *visibility*, $V(\mathbf{u}, \nu)$, which is the angular Fourier mode of the intensity fluctuations on the sky measured by a baseline \mathbf{u} in a frequency channel ν . In the flat-sky approximation, the complex visibility can be written simply as the 2D Fourier transform of the beam-modulated sky intensity:

$$V(\mathbf{u}, \nu) = \int d^2\theta A(\theta, \nu) T(\theta, \nu) e^{-2\pi i \mathbf{u} \cdot \theta}, \quad (1)$$

where θ denotes the angular sky position, $T(\theta, \nu)$ is the sky temperature, $A(\theta, \nu)$ is the primary beam pattern, which determines the observed area of the sky, and the baseline vector \mathbf{u} is on the uv -plane perpendicular to the direction of the incoming signals, and measures the separation between the two antenna pairs in units of wavelength, i.e., $\mathbf{u} = \nu/c \Delta \mathbf{r}$. As the physical separation $\Delta \mathbf{r}$ of each pair of antennas is assumed to be fixed, each frequency channel thus gives an independent set of visibility samplings in the uv -plane with frequency-dependent baseline vectors.

In practice, especially for 21-cm tomography, one discretizes the 3D sky into voxels so as to efficiently estimate

the intensity distribution $T(\theta, \nu)$ and associated power spectrum using FFT-based algorithms. If we assume n_\perp pixels per 2D image (real or Fourier space) with even spacings in the θ_x - and θ_y -directions (i.e. two transverse directions), and an instrument with n_\parallel frequency channels of an equal spectral resolution $\Delta\nu$ to probe the sky signal in a radial direction, then we have a 3D uniform Cartesian grid of size $n_\perp n_\parallel$. Throughout the paper, the symbols \perp and \parallel denote the two transverse directions and the radial direction, respectively.

Following Sutter et al. (2012), the radio interferometric measurement equation of Eq. 1 at the frequency ν_i , for $i = 1, 2, \dots, n_\parallel$, can be recast in a discrete operator formalism by

$$\mathbf{y}_\perp[\nu_i] = \mathbf{I}[\nu_i] \mathbf{F}_\perp \mathbf{A}[\nu_i] \mathbf{F}_\perp^{-1} \tilde{\boldsymbol{\psi}}_\perp[\nu_i] + \mathbf{n}_\perp[\nu_i], \quad (2)$$

where $\mathbf{y}_\perp[\nu_i] = (y[u_1, \nu_i], \dots, y[u_{n_\perp}, \nu_i])^T$ and $\tilde{\boldsymbol{\psi}}_\perp[\nu_i] = (\tilde{\psi}[u_1, \nu_i], \dots, \tilde{\psi}[u_{n_\perp}, \nu_i])^T$ denote, respectively, the discretized visibilities in the uv -plane and the 2D Fourier “image” of the true angular sky ($\boldsymbol{\psi}_\perp[\nu_i]$) of 21-cm and foregrounds at the i -th channel. The vector $\mathbf{n}_\perp[\nu_i] = (n[u_1, \nu_i], \dots, n[u_{n_\perp}, \nu_i])^T$ denotes the instrumental noise at the i -th channel for all the uv points. Hereafter, bold, upper and lower case letters denote a matrix and a vector, respectively, and the symbol \sim is used for Fourier components. \mathbf{F}_\perp represents the 2D Fourier transform operator performed in the transverse direction, converting from the real-space domain into the uv -space domain (of course, \mathbf{F}_\perp^{-1} is its inverse), and $\mathbf{A}[\nu_i]$ represents the primary beam pattern of the interferometer measured at the frequency ν_i and implicitly is a $n_\perp \times n_\perp$ diagonal matrix. The operator $\mathbf{I}[\nu_i]$ specifies the visibility sampling function determined by the distribution of baselines and can also be represented by a $n_\perp \times n_\perp$ diagonal matrix that has 1’s and 0’s on the diagonal to indicate whether a given pixel in uv -space has been observed or not.

It is convenient to introduce a vector form for numerical operations. By stacking each vector $\tilde{\boldsymbol{\psi}}_\perp[\nu_i]$ as $\tilde{\boldsymbol{\psi}}_\perp = [\tilde{\boldsymbol{\psi}}_\perp^T[\nu_1], \dots, \tilde{\boldsymbol{\psi}}_\perp^T[\nu_{n_\parallel}]]^T$, Eq. 2 can account for all the frequency channels when rewritten as

$$\mathbf{y} = \boldsymbol{\Phi} \tilde{\boldsymbol{\psi}}_\perp + \mathbf{n}, \quad \boldsymbol{\Phi} = \begin{pmatrix} \boldsymbol{\Phi}_1 & \mathbf{0} \\ & \ddots \\ \mathbf{0} & \boldsymbol{\Phi}_{n_\parallel} \end{pmatrix}, \quad (3)$$

$$\text{with } \boldsymbol{\Phi}_i = \mathbf{I}[\nu_i] \mathbf{F}_\perp \mathbf{A}[\nu_i] \mathbf{F}_\perp^{-1},$$

where $\tilde{\boldsymbol{\psi}}_\perp$ is a vector with length $n_\perp n_\parallel$ and $\boldsymbol{\Phi}$ is a block diagonal matrix, while $\mathbf{y} = [\mathbf{y}_\perp^T[\nu_1], \dots, \mathbf{y}_\perp^T[\nu_{n_\parallel}]]^T$, and $\mathbf{n} = [\mathbf{n}_\perp^T[\nu_1], \dots, \mathbf{n}_\perp^T[\nu_{n_\parallel}]]^T$, each collect vectors for all frequencies into an $n_\perp n_\parallel$ -element vector.

Notice that Eq. 3 is presented on a uniform grid in order to enable the FFT operation. Realistic observations from radio interferometers do not sample visibility data on a uniform grid in uv -space. Therefore, analyzing the data on a uniform grid requires an additional process, “gridding”, to map raw visibilities from specified uv points to grids with even spacing. However, the gridding process can distort the signal estimation, especially for small 21-cm signals. Alternatively, one can use the Non Uniform Discrete Fourier Transform (NuDFT) (Fessler

& Sutton 2003) to directly calculate visibility data at continuous coordinates so as to avoid “gridding” effects. We leave an investigation of the gridding effects and a possible solution by NuDFT to future work.

2.2. Linear mixture model

The 21-cm signal along the line-of-sight for each individual pixel is not a smooth function of frequency, while most astrophysical contaminants should have smooth power-law spectral structures. If we assume that the total foreground emission contributing to the observed data in a given frequency channel can be expressed as a linear superposition of a few components and the spectral behavior of each component does not vary across the sky, the true sky temperature in pixel θ_k , for $k = 1, 2, \dots, n_\perp$, in the i -th channel then can be modelled as

$$\psi[\nu_i, \theta_k] = s[\nu_i, \theta_k] + \sum_{j=1}^{n_c} M_{ij} f_j[\theta_k], \quad (4)$$

where ν_i is the frequency, θ_k is the pixel position, n_c the assumed total number of foreground components, $s[\nu_i, \theta_k]$ the 21-cm signal and $f_j[\theta_k]$ the j -th foreground component. M_{ij} reflects the contribution of the j -th foreground component to the i -th frequency channel and is assumed to be constant and independent of θ_k , and thus the matrix \mathbf{M} is commonly called the “mixing matrix” with dimensions of $n_\parallel \times n_c$. Each column of \mathbf{M} encodes the frequency-dependence of each foreground component.

As the Fourier modes of the isotropic 21-cm signal $s(\nu, \boldsymbol{\theta})$ are expected to be mutually independent, expressing its temperature in the Fourier coordinates (η, \mathbf{u}) is particularly useful. By taking the 3D Fourier transform along the frequency axis and two angular directions, one has

$$\tilde{s}(\eta, \mathbf{u}) = \int s(\nu, \boldsymbol{\theta}) e^{-2\pi i (\nu\eta + \mathbf{u}\cdot\boldsymbol{\theta})} d\boldsymbol{\theta} d\nu, \quad (5)$$

where η is the Fourier dual of the frequency variable ν . Note that here we perform the “true” Fourier transform of the sky temperature along the line-of-sight direction that is perpendicular to the uv -plane, rather than a “delay transform”. The delay transform is the Fourier transform in frequency of the spectrum measured by a single baseline, which varies with frequency (i.e., \mathbf{u} is a function of ν in Eq. 5) and which leads to the “mode-mixing” phenomenon (see Liu et al. (2014a) for details).

The foreground components can also be expressed in terms of their 2D Fourier modes. For the j -th component, one has

$$\tilde{f}_j(\mathbf{u}) = \int f_j(\boldsymbol{\theta}) e^{-2\pi i \mathbf{u}\cdot\boldsymbol{\theta}} d\boldsymbol{\theta} \quad (j = 1, \dots, n_c). \quad (6)$$

According to the above equations, by performing the angular Fourier transform on Eq. 4 and expressing the 21-cm signal in terms of its 3D Fourier modes, Eq. 4 can be rewritten in a discretized version as

$$\tilde{\psi}[\nu_i, u_k] = \sum_{m=1}^{n_\parallel} F_\parallel^{-1}[i, m] \tilde{s}[\eta_m, u_k] + \sum_{j=1}^{n_c} M_{ij} \tilde{f}_j[u_k], \quad (7)$$

where i denotes the frequency channel number, k denotes pixel position in the uv -plane and we have introduced the operator (matrix) $\mathbf{F}_\parallel^{-1}$, with its elements given

by Fourier coefficients $F_\parallel^{-1}[i, m]$ ($i = 1, \dots, n_\parallel; m = 1, \dots, n_\parallel$), to represent the 1-D inverse Fourier transform converting the η -domain to the frequency domain.

Using matrix multiplications, we can rewrite the linear system above as the matrix equation:

$$\tilde{\Psi} = \mathbf{F}_\parallel^{-1} \tilde{\mathbf{S}} + \mathbf{M} \tilde{\mathbf{F}}, \quad (8)$$

where $\tilde{\psi}[\nu_i, u_k]$ is the (i, k) -th entry of the matrix $\tilde{\Psi}$, $\tilde{s}[\eta_m, u_k]$ is the (m, k) -th entry of the matrix $\tilde{\mathbf{S}}$, and $\tilde{f}_j[u_k]$ the (j, k) -th entry of the matrix $\tilde{\mathbf{F}}$. Applying a “vectorization” operator ($\text{vec}[\cdot]$) that converts the matrix into a column vector by stacking the columns into a long column vector, one can express $\tilde{\psi}_\perp$ in Eq. 3 with $\tilde{\Psi}^T$ by

$$\tilde{\psi}_\perp = \text{vec}[\tilde{\Psi}^T] = \text{vec}[(\mathbf{F}_\parallel^{-1} \tilde{\mathbf{S}})^T] + \text{vec}[(\mathbf{M} \tilde{\mathbf{F}})^T]. \quad (9)$$

Substituting Eq. 9 into 3, and using the identity $\text{vec}[\mathbf{P}\mathbf{Q}] = (\mathbf{Q}^T \otimes \mathbf{I}_m) \text{vec}[\mathbf{P}]$ for any matrices $\mathbf{P}_{m \times n}$ and $\mathbf{Q}_{n \times p}$, where \otimes denotes the Kronecker product and \mathbf{I}_m is an $(m \times m)$ identity matrix, yields

$$\tilde{\psi}_\perp = (\mathbf{B} \otimes \mathbf{I}_{n_\perp}) \mathbf{x}, \quad \mathbf{B} = (\mathbf{F}_\parallel^{-1} \mathbf{M}), \quad (10)$$

where $\mathbf{x} = (\tilde{\mathbf{s}}^T, \tilde{\mathbf{f}}^T)^T$ with $\tilde{\mathbf{s}} = \text{vec}[\tilde{\mathbf{S}}^T]$, $\tilde{\mathbf{f}} = \text{vec}[\tilde{\mathbf{F}}^T]$, \mathbf{I}_{n_\perp} is the $n_\perp \times n_\perp$ identity matrix, and \mathbf{B} is a partitioned matrix with dimensions of $n_\parallel \times (n_\parallel + n_c)$. Explicitly, $\tilde{\mathbf{s}}$ and $\tilde{\mathbf{f}}$ are vectors of length $n_\perp n_\parallel$ and $n_\perp n_c$, respectively. They are constructed by stacking $\tilde{s}[\eta_m]$ and \tilde{f}_j as $\tilde{\mathbf{s}} = (\tilde{s}^T[\eta_1], \dots, \tilde{s}^T[\eta_{n_\parallel}])^T$ and $\tilde{\mathbf{f}} = (\tilde{f}_1^T, \dots, \tilde{f}_{n_c}^T)^T$, where the vectors $\tilde{s}[\eta_m]$ and \tilde{f}_j , each with length n_\perp , denote an image of the 3D Fourier modes of the 21-cm signal at η_m and an image of the 2D Fourier modes of the j -th foreground component, respectively, i.e., $\tilde{s}[\eta_m] = (\tilde{s}[u_1, \eta_m], \dots, \tilde{s}[u_{n_\perp}, \eta_m])$ and $\tilde{f}_j = (\tilde{f}_j[u_1], \dots, \tilde{f}_j[u_{n_\perp}])$.

By inserting Eq. 10 into Eq. 3, the measurement equation finally can be expressed as

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \quad \mathbf{H} = \Phi(\mathbf{B} \otimes \mathbf{I}_{n_\perp}). \quad (11)$$

2.3. ICA assumption

The Independent Component Analysis (ICA) assumption is that the data can be considered as a linear mixture of a set of statistically mutually independent components. The cosmological 21-cm signal is expected to be well approximated by an isotropic Gaussian random field and uncorrelated with foregrounds. If we also assume that the diffuse foregrounds consist of several statistically Independent Components (ICs), each of them also being an isotropic Gaussian random field with zero mean, as defined in Eq. 10, the covariance matrix of \mathbf{x} then becomes a diagonal matrix, namely

$$\mathbf{C} = \langle \mathbf{x}\mathbf{x}^\dagger \rangle = \begin{pmatrix} \langle \tilde{\mathbf{s}}\tilde{\mathbf{s}}^\dagger \rangle & \mathbf{0} \\ \mathbf{0} & \langle \tilde{\mathbf{f}}\tilde{\mathbf{f}}^\dagger \rangle \end{pmatrix}. \quad (12)$$

The unknown diagonal matrices $\langle \tilde{\mathbf{s}}\tilde{\mathbf{s}}^\dagger \rangle$ and $\langle \tilde{\mathbf{f}}\tilde{\mathbf{f}}^\dagger \rangle$ are to be estimated from the observed data. They determine the 3D power spectrum of the HI signal and the angular power spectra of the ICs, defined through

$$\langle \tilde{s}[u_k, \eta_m] \tilde{s}^*[u_{k'}, \eta_{m'}] \rangle = P_{\text{HI}}(u_k, \eta_m) \delta_{kk'} \delta_{mm'} \quad (13)$$

$$\langle \tilde{f}_j(u_k) \tilde{f}_j^*(u_{k'}) \rangle = \delta_{jj'} \delta_{kk'} C_f^j(\ell = 2\pi|\mathbf{u}|) \quad (14)$$

where $P_{\text{HI}}(u_k, \eta_m)$ denotes the 3D power spectrum of the 21-cm signal as a function of \mathbf{u} and η , and $C_f^j(\ell)$ denotes the angular power spectrum of the j -th foreground component as a function of multipole ℓ . The relation $\ell = 2\pi|\mathbf{u}|$, where \mathbf{u} is determined by the pixel position u_k , has been used in the flat-sky approximation. Notice that physical sources such as the Galactic synchrotron and free-free emissions actually do have a non-zero cross-correlation and their spatial distributions clearly appear to be non-isotropic. Nevertheless, as we will demonstrate, the ICA assumption can be regarded as an effective decomposition of sources and does not appear to affect our ability to remove foregrounds.

2.3.1. Noise

We also assume that the instrument noise is an uncorrelated Gaussian distribution and can be obtained from a reasonable noise model through

$$\langle \tilde{n}(u_k, \nu_i) \tilde{n}^*(u_{k'}, \nu_{i'}) \rangle = P_N(u_k, \nu_i) \delta_{kk'} \delta_{ii'}. \quad (15)$$

The covariance matrix thus has the form of a known block diagonal matrix (with \mathbf{n} defined in Eq. 3), where each block is also diagonal,

$$\mathbf{N} = \langle \mathbf{n} \mathbf{n}^\dagger \rangle = \begin{pmatrix} \mathbf{N}_1 & & \mathbf{0} \\ & \ddots & \\ \mathbf{0} & & \mathbf{N}_{n_\parallel} \end{pmatrix}. \quad (16)$$

Here \mathbf{N}_i denotes the $n_\perp \times n_\perp$ covariance matrix of the noise map in the uv -plane at the i -th frequency channel, which is also diagonal.

2.4. The EM algorithm

The goal of our analysis is to identify and separate the components from visibility data that contains mixtures of foregrounds and signal. Using the Bayesian framework proposed by Snoussi et al. (2002) for CMB data, we will show that, without any assumption on priors, by performing a semi-blind independent component analysis for which only the noise covariance matrix \mathbf{N} is well known, one can successfully separate the components by jointly estimating the covariance matrix \mathbf{C} and the mixing matrix \mathbf{M} , and can accurately estimate the power spectrum of the HI signal.

In Bayesian inference, information about unknown parameters $\boldsymbol{\theta}$ that we want to estimate from the data \mathbf{y} is expressed in the form of a posterior probability distribution. Using Bayes' theorem, it can be computed as $p(\boldsymbol{\theta}|\mathbf{y}) \propto \mathcal{L}(\mathbf{y}|\boldsymbol{\theta})p(\boldsymbol{\theta})$ where $\mathcal{L}(\mathbf{y}|\boldsymbol{\theta})$ is the likelihood and $p(\boldsymbol{\theta})$ is the prior distribution of $\boldsymbol{\theta}$; $\boldsymbol{\theta} = (\mathbf{C}, \mathbf{M})$ in our case. If we assume flat priors for the mixing matrix and the power spectra of all sources, and assume they are uncorrelated and independent, the prior then reduces to $P(\boldsymbol{\theta}) = P(\mathbf{C})P(\mathbf{M}) \propto 1$. Thus, exploring the observed-data posterior $p(\mathbf{y}|\boldsymbol{\theta})$ is equivalent to exploring the likelihood. Following Snoussi et al. (2002), given the data model in Eq. 11, $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}$, the mixing matrix \mathbf{M} and the covariance matrix \mathbf{C} can be estimated by maximizing the likelihood function. For independent and Gaussian sources, the likelihood can be expressed as

$$-2 \ln \mathcal{L}(\mathbf{y}|\boldsymbol{\theta}) \propto \ln |\mathbf{H}\mathbf{C}\mathbf{H}^\dagger + \mathbf{N}| + \text{Tr} [(\mathbf{H}\mathbf{C}\mathbf{H}^\dagger + \mathbf{N})^{-1} \mathbf{y}\mathbf{y}^\dagger]. \quad (17)$$

Unfortunately, in general, such a likelihood evaluation is computationally intractable when applied to large datasets with realistic interferometric observations. However, in the analogous case of CMB imaging observations[†], this likelihood can be approximated by the SMICA-likelihood (Delabrouille et al. 2003) to measure a “spectral mismatch” in the Fourier domain between the empirical covariance matrices of the data and their ensemble averages, which depend on the estimated parameters. In practice, minimization of the spectral mismatch is equivalent to maximizing the likelihood and can be achieved with the Expectation-Maximization (EM) algorithm. The EM algorithm has been applied to estimate the CMB power spectrum and reconstruct the CMB map from multi-frequency microwave maps (Delabrouille et al. 2003). The EM algorithm is an elaborate technique to find the maximum-likelihood estimate of the parameters when directly evaluating the likelihood function is intractable. The calculation is simplified by assuming the existence of additional but missing parameters. It is an iterative algorithm to repeatedly solve a tractable complete-data problem instead of solving a difficult incomplete-data problem.

Here we briefly summarize the EM algorithm when applied to our model. The likelihood $p(\mathbf{y}|\boldsymbol{\theta})$ first can be obtained by marginalizing the joint distribution $p(\mathbf{y}, \mathbf{x}|\boldsymbol{\theta})$ over the missing data \mathbf{x} as

$$p(\mathbf{y}|\boldsymbol{\theta}) = \int p(\mathbf{y}, \mathbf{x}|\boldsymbol{\theta}) d\mathbf{x}. \quad (18)$$

The key idea is that the EM algorithm does not maximize $p(\mathbf{y}|\boldsymbol{\theta})$ directly; instead, it maximizes the so-called EM-functional as follows:

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^n) = \mathbb{E} [\ln p(\mathbf{y}, \mathbf{x}|\boldsymbol{\theta}) | \mathbf{y}, \boldsymbol{\theta}^n] = \int \ln p(\mathbf{y}, \mathbf{x}|\boldsymbol{\theta}) p(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta}^n) d\mathbf{x}, \quad (19)$$

which is the expected value of the complete-data log-likelihood with respect to the missing data \mathbf{x} given the observed data \mathbf{y} and the current parameter estimates $\boldsymbol{\theta}^n$. The evaluation of this expectation is called the *Expectation (E)-step* of the algorithm.

In our model, the prior distribution of the complex Fourier modes \mathbf{x} is assumed to be Gaussian with zero mean,

$$p(\mathbf{x}|\mathbf{C}) \propto \exp \left(-\frac{1}{2} \mathbf{x}^\dagger \mathbf{C}^{-1} \mathbf{x} \right), \quad (20)$$

where the diagonal covariance matrix \mathbf{C} is defined in Eq. 12. Within a Bayesian framework, the joint posterior distribution $p(\mathbf{y}, \mathbf{x}|\boldsymbol{\theta})$ can be expressed by $p(\mathbf{y}, \mathbf{x}|\boldsymbol{\theta}) = p(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})p(\mathbf{x}|\boldsymbol{\theta})$, yielding

$$-2 \ln p(\mathbf{y}, \mathbf{x}|\boldsymbol{\theta}) = \ln |\mathbf{N}| + (\mathbf{y} - \mathbf{H}\mathbf{x})^\dagger \mathbf{N}^{-1} (\mathbf{y} - \mathbf{H}\mathbf{x}) + \ln |\mathbf{C}| + \mathbf{x}^\dagger \mathbf{C}^{-1} \mathbf{x} + \text{cst}. \quad (21)$$

Using Bayes' rule, the conditional probability distribution function $p(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta}^n)$ for the signal given the data is also the Gaussian,

$$p(\mathbf{x}|\mathbf{y}, \boldsymbol{\theta}^n) \propto \exp \left(-\frac{1}{2} (\mathbf{x} - \hat{\mathbf{x}})^\dagger \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \hat{\mathbf{x}}) \right), \quad (22)$$

[†] The convolution with a frequency-dependent beam response in real space becomes a simple product in Fourier space.

where

$$\hat{\mathbf{x}} = [\mathbf{H}^\dagger \mathbf{N}^{-1} \mathbf{H} + \mathbf{C}^{-1}]^{-1} \mathbf{H}^\dagger \mathbf{N}^{-1} \mathbf{y}, \quad (23)$$

$$\boldsymbol{\Sigma} = (\mathbf{H}^\dagger \mathbf{N}^{-1} \mathbf{H} + \mathbf{C}^{-1})^{-1}, \quad (24)$$

where $\hat{\mathbf{x}}$ is the so-called Wiener-Filtered (WF) map and $\boldsymbol{\Sigma}$ is the corresponding covariance $\boldsymbol{\Sigma} = \langle (\mathbf{x} - \hat{\mathbf{x}})(\mathbf{x} - \hat{\mathbf{x}})^\dagger \rangle$. The solution for $\hat{\mathbf{x}}$ is the general map-making problem in cosmology. This WF map can be computed efficiently by implementing a preconditioned conjugate-gradient method that allows one to iteratively reach the solution in a tractable amount of computation time.

Using the above Eqs. 21 and 22, we integrate out \mathbf{x} to derive the expression of Eq. 19 given by

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^n) = \text{cst.} - \frac{1}{2} \ln |\mathbf{N}| - \frac{1}{2} \ln |\mathbf{C}| - \frac{1}{2} \text{Tr} \left[\mathbf{C}^{-1} \hat{\mathbf{R}}_{xx}^\dagger \right] - \frac{1}{2} \text{Tr} \left[\mathbf{N}^{-1} \left(\hat{\mathbf{R}}_{yy} - \mathbf{H} \hat{\mathbf{R}}_{yx}^\dagger - \hat{\mathbf{R}}_{yx} \mathbf{H}^\dagger + \mathbf{H} \hat{\mathbf{R}}_{xx} \mathbf{H}^\dagger \right) \right] \quad (25)$$

where

$$\hat{\mathbf{R}}_{yy} = \mathbf{y} \mathbf{y}^\dagger, \quad \hat{\mathbf{R}}_{yx} = \mathbf{y} \hat{\mathbf{x}}^\dagger, \quad \hat{\mathbf{R}}_{xx} = \boldsymbol{\Sigma} + \hat{\mathbf{x}} \hat{\mathbf{x}}^\dagger. \quad (26)$$

Notice here $\hat{\mathbf{R}}_{yy}$, $\hat{\mathbf{R}}_{yx}$ and $\hat{\mathbf{R}}_{xx}$ only depend on $\boldsymbol{\theta}^n$ rather than $\boldsymbol{\theta}$.

The second step, called the *Maximization (M)-step*, updates the parameters by maximizing the expectation we computed in the previous E-step. These two steps are repeated as necessary. This procedure is guaranteed to increase the likelihood $p(\mathbf{y}|\boldsymbol{\theta})$ monotonically with successive iterations. In order to obtain the parameter $\boldsymbol{\theta}^{n+1}$ at iteration $n+1$, we solve the gradient equation with respect to \mathbf{M} and \mathbf{C} to maximize the functional Q . To do so, let us first introduce some notation that allows us to express the derivative simply.

Let $\hat{\mathbf{x}}$ and $\boldsymbol{\Sigma}$ be partitioned as follows:

$$\hat{\mathbf{x}} = \left(\hat{\mathbf{s}}^T, \hat{\mathbf{f}}^T \right)^T \quad (27)$$

with

$$\hat{\mathbf{s}} = \left(\hat{\mathbf{s}}_1^T, \dots, \hat{\mathbf{s}}_{n_\perp}^T \right)^T, \quad \hat{\mathbf{f}} = \left(\hat{\mathbf{f}}_1^T, \dots, \hat{\mathbf{f}}_{n_c}^T \right)^T. \quad (28)$$

Here $\hat{\mathbf{x}}$ is equally divided into $(n_\parallel + n_c)$ subvectors, each with length n_\perp , $\hat{\mathbf{s}}_\perp[\eta_i]$ denotes the WF map of the 21-cm signal at η_i , and $\hat{\mathbf{f}}_j$ denotes the WF map of the j -th foreground component. Similarly

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}^{\tilde{s}\tilde{s}} & \boldsymbol{\Sigma}^{\tilde{s}\tilde{f}} \\ \boldsymbol{\Sigma}^{\tilde{s}\tilde{f}\dagger} & \boldsymbol{\Sigma}^{\tilde{f}\tilde{f}} \end{pmatrix} \quad (29)$$

with

$$\boldsymbol{\Sigma}^{\tilde{s}\tilde{s}} = \begin{pmatrix} \boldsymbol{\Sigma}_{11}^{\tilde{s}\tilde{s}} & \dots & \boldsymbol{\Sigma}_{1n_\perp}^{\tilde{s}\tilde{s}} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{n_\perp 1}^{\tilde{s}\tilde{s}} & \dots & \boldsymbol{\Sigma}_{n_\perp n_\perp}^{\tilde{s}\tilde{s}} \end{pmatrix}, \quad \boldsymbol{\Sigma}^{\tilde{s}\tilde{f}} = \begin{pmatrix} \boldsymbol{\Sigma}_{11}^{\tilde{s}\tilde{f}} & \dots & \boldsymbol{\Sigma}_{1n_c}^{\tilde{s}\tilde{f}} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{n_\perp 1}^{\tilde{s}\tilde{f}} & \dots & \boldsymbol{\Sigma}_{n_\perp n_c}^{\tilde{s}\tilde{f}} \end{pmatrix}$$

$$\boldsymbol{\Sigma}^{\tilde{f}\tilde{f}} = \begin{pmatrix} \boldsymbol{\Sigma}_{11}^{\tilde{f}\tilde{f}} & \dots & \boldsymbol{\Sigma}_{1n_c}^{\tilde{f}\tilde{f}} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Sigma}_{n_c 1}^{\tilde{f}\tilde{f}} & \dots & \boldsymbol{\Sigma}_{n_c n_c}^{\tilde{f}\tilde{f}} \end{pmatrix}, \quad (30)$$

where

$$\boldsymbol{\Sigma}_{ij}^{\tilde{s}\tilde{s}} = \left\langle (\tilde{\mathbf{s}}[\eta_i] - \hat{\mathbf{s}}[\eta_i])(\tilde{\mathbf{s}}[\eta_j] - \hat{\mathbf{s}}[\eta_j])^\dagger \right\rangle, \quad (31)$$

$$\boldsymbol{\Sigma}_{ij}^{\tilde{s}\tilde{f}} = \left\langle (\tilde{\mathbf{s}}[\eta_i] - \hat{\mathbf{s}}[\eta_i])(\tilde{\mathbf{f}}[\eta_j] - \hat{\mathbf{f}}[\eta_j])^\dagger \right\rangle, \quad (32)$$

$$\boldsymbol{\Sigma}_{ij}^{\tilde{f}\tilde{f}} = \left\langle (\tilde{\mathbf{f}}[\eta_i] - \hat{\mathbf{f}}[\eta_i])(\tilde{\mathbf{f}}[\eta_j] - \hat{\mathbf{f}}[\eta_j])^\dagger \right\rangle, \quad (33)$$

each with dimensions of $n_\perp \times n_\perp$.

The gradient equation then reads:

$$\frac{\partial Q}{\partial \mathbf{C}} = -(\mathbf{C}^{-1}) + \mathbf{C}^{-1} \hat{\mathbf{R}}_{xx} \mathbf{C}^{-1} = 0 \quad (34)$$

$$\frac{\partial Q}{\partial M_{ij}} = \text{Tr} \left[\mathbf{N}_i^{-1} \left(\Phi_i \hat{\mathbf{f}}_j \mathbf{y}_i^\dagger + (\Phi_i \hat{\mathbf{f}}_j \mathbf{y}_i^\dagger)^\dagger \right) \right] - \frac{\partial \text{Tr} [\mathbf{N}^{-1} \mathbf{H} \hat{\mathbf{R}}_{xx} \mathbf{H}^\dagger]}{\partial M_{ij}} = 0 \quad (35)$$

where the derivative of the last term can be well approximated by

$$\frac{\partial \text{Tr} [\mathbf{N}^{-1} \mathbf{H} \hat{\mathbf{R}}_{xx} \mathbf{H}^\dagger]}{\partial M_{ij}} \approx \quad (36)$$

$$\sum_{k=1}^{n_c} M_{ik} \text{Tr} \left[\mathbf{N}_i^{-1} \Phi_i \left(\boldsymbol{\Sigma}_{kj}^{\tilde{f}\tilde{f}} + \hat{\mathbf{f}}_k \hat{\mathbf{f}}_j^\dagger + (\boldsymbol{\Sigma}_{kj}^{\tilde{f}\tilde{f}} + \hat{\mathbf{f}}_k \hat{\mathbf{f}}_j^\dagger)^\dagger \right) \Phi_i^\dagger \right]$$

where $\Phi_i = \mathbf{I}[\nu_i] \mathbf{F}_\perp \mathbf{A}[\nu_i] \mathbf{F}_\perp^{-1}$ as defined in Eq. 2, and we have neglected the terms that include $\boldsymbol{\Sigma}^{\tilde{s}\tilde{f}}$ and $\boldsymbol{\Sigma}^{\tilde{s}\tilde{s}}$, although the complete expression can be obtained straightforwardly. This is an excellent approximation for 21-cm foregrounds since they are several orders of magnitude brighter than the 21-cm signal[†]. By making this approximation, we are able to dramatically speed up the computation. According to Eqs. 34, 35 and 37, one can establish the *update scheme* as follows:

$$\mathbf{M}_{i,:} \leftarrow \left(\rho_i^{\tilde{f}y}[1], \dots, \rho_i^{\tilde{f}y}[n_c] \right) \begin{pmatrix} \rho_i^{\tilde{f}\tilde{f}}[1,1] & \dots & \rho_i^{\tilde{f}\tilde{f}}[1,n_c] \\ \vdots & \ddots & \vdots \\ \rho_i^{\tilde{f}\tilde{f}}[n_c,1] & \dots & \rho_i^{\tilde{f}\tilde{f}}[n_c,n_c] \end{pmatrix}^{-1} \quad (37)$$

$$\mathbf{C}_{ii} \leftarrow (\hat{\mathbf{R}}_{xx})_{ii} \quad (38)$$

where $\mathbf{M}_{i,:}$ denotes the i -th row of the mixing matrix, and

$$\rho_i^{\tilde{f}\tilde{f}}[k,j] = \text{Re} \left\{ \text{Tr} \left[\mathbf{N}_i^{-1} \Phi_i (\boldsymbol{\Sigma}_{kj}^{\tilde{f}\tilde{f}} + \hat{\mathbf{f}}_k \hat{\mathbf{f}}_j^\dagger) \Phi_i^\dagger \right] \right\} \quad (39)$$

$$\rho_i^{\tilde{f}y}[j] = \text{Re} \left\{ \text{Tr} \left[\mathbf{N}_i^{-1} \Phi_i \hat{\mathbf{f}}_j \mathbf{y}_i^\dagger \right] \right\}. \quad (40)$$

One can verify that, by setting Φ_i as an identity matrix, for $i = 1, 2, \dots, n_\parallel$, such that no instrumental effects are present, this is in complete agreement with the structure of the expression for the mixing matrix in the literature (Snoussi et al. 2002; Delabrouille et al. 2003).

The problem is evaluating the matrix $\boldsymbol{\Sigma}$ from its inverse as given by Eq. 24 in order to obtain the covariance matrix \mathbf{C} from Eq. 38. Since the matrix $\boldsymbol{\Sigma}$ has the dimensions of $n_\perp(n_\parallel + n_c) \times n_\perp(n_\parallel + n_c)$, which is typically of order $10^6 \times 10^6$ in a 21-cm survey, such matrix inversion is completely prohibitive computationally.

[†] Notice that, in the case of CMB, the evaluation of the mixing matrix not only depends on CMB foregrounds but also on the CMB signal itself since they are roughly comparable.

However, as the power spectrum is only determined by the diagonal components of $\mathbf{\Sigma}$, not by the off-diagonals, by using the same trick presented in the Gibbs sampling techniques (Wandelt et al. 2004; Jewell et al. 2004) for CMB data analysis, one can verify that the ensemble-averaged solution for the following linear equation has the desired covariance. By solving

$$[\mathbf{H}^\dagger \mathbf{N}^{-1} \mathbf{H} + \mathbf{C}^{-1}] \mathbf{z} = \mathbf{H}^\dagger \mathbf{N}^{-1/2} \mathbf{z}_1 + \mathbf{C}^{-1/2} \mathbf{z}_2 \quad (41)$$

where the real vectors \mathbf{z}_1 and \mathbf{z}_2 are of length $n_\perp n_\parallel$ and $n_\perp(n_c + n_\parallel)$, respectively, with elements drawn from a standard normal distribution, one can find that $\mathbf{\Sigma} = \langle \mathbf{z} \mathbf{z}^\dagger \rangle$ where $\langle \cdot \rangle$ represents the average over all solutions for a large number of realizations of \mathbf{z}_1 and \mathbf{z}_2 . This algorithm can rapidly provide accurate approximate solutions in massively parallel computers. Therefore one can obtain the diagonal components of $\hat{\mathbf{R}}_{\mathbf{xx}}$ in Eq. 38 efficiently by $(\hat{\mathbf{R}}_{\mathbf{xx}})_{ii} = \langle |z_i|^2 \rangle + |\hat{x}_i|^2$ where i runs over all elements. In the same way as in $\hat{\mathbf{x}}$, the vector \mathbf{z} with length $(n_\parallel + n_c)n_\perp$ can also be equally split into $(n_\parallel + n_c)$ subvectors, each with length n_\perp , as $\mathbf{z} = \left((\mathbf{z}_1^{\tilde{s}})^T, \dots, (\mathbf{z}_{n_\parallel}^{\tilde{s}})^T, (\mathbf{z}_1^{\tilde{f}})^T, \dots, (\mathbf{z}_{n_c}^{\tilde{f}})^T \right)^T$, in correspondence with contributions from the 21-cm signal, denoted by $\mathbf{z}_i^{\tilde{s}}$ for $i = 1, \dots, n_\perp$, and the ICs, denoted by $\mathbf{z}_i^{\tilde{f}}$ for $i = 1, \dots, n_c$. Note that, in general the noise could be correlated, resulting in non-zero off-diagonal components in the noise covariance matrix, \mathbf{N} . Solving Eq. 41 will require multiplication by a dense $n_\perp n_\parallel \times n_\perp n_\parallel$ inverse noise covariance matrix, with a computational scaling of $\mathcal{O}(n_\perp^2 n_\parallel^2)$. In the most general case, inverting \mathbf{N} is a significant computational challenge. However, the inverse noise covariance matrix needs only to be computed once and stored, and then can be used many times to compute Eqs. 39, 40 and 41.

Using the above mentioned trick, the trace term in Eq. 39 can be computed merely by the ensemble averaged dot products of two vectors as

$$\text{Tr} \left[\mathbf{N}_i^{-1} \Phi_i (\mathbf{\Sigma}_{k_j}^{\tilde{f}\tilde{f}} + \hat{\mathbf{f}}_k \hat{\mathbf{f}}_k^\dagger) \Phi_i^\dagger \right] = \left\langle (\Phi_i \mathbf{z}_j^{\tilde{f}})^\dagger (\mathbf{N}_i^{-1} \Phi_i \mathbf{z}_k^{\tilde{f}}) \right\rangle + (\Phi_i \hat{\mathbf{f}}_j)^\dagger \mathbf{N}_i^{-1} (\Phi_i \hat{\mathbf{f}}_k), \quad (42)$$

where the vectors $\mathbf{z}_j^{\tilde{f}}$ and $\mathbf{z}_k^{\tilde{f}}$ represent the subvectors of \mathbf{z} solved by Eq. 41. The above equation applies also to the case of correlated noise.

Moreover, in each EM iteration, we fix the magnitude of each column of \mathbf{M} to unity and adjust the corresponding power spectra of the ICs accordingly, similar to Delabrouille et al. (2003), to break the degeneracy and keep the scale-invariant product $\sum_j M_{ij} M_{i'j} \langle \tilde{f}_j[u_k] \tilde{f}_j^*[u_k] \rangle = \sum_j M_{ij} M_{i'j} C_f^j[u_k]$ unchanged for an arbitrary pixel u_k at any frequency channels i and i' , as seen from Eq. 7.

After updating \mathbf{M} and \mathbf{C} , since the quantities, $\ell^2 C_\ell^i$ for the ICs and $k_\parallel k_\perp^2 P_{\text{HI}}$ for the HI signal are expected to vary more slowly than C_ℓ^i and P_{HI} themselves, it is more appropriate to perform bin averages after each iteration and update the associated elements in \mathbf{C} , while ensuring the number of observations be smaller than the number of the estimated parameters in \mathbf{C} and \mathbf{M} . The simulation shows that using band-averaged spectra can

significantly increase the convergence speed and obtain much more stable spectra with high accuracy in iterations. According to Eq 13, the bin-averages then yield the following estimates:

$$C_f^j(\mathbf{u}) \leftarrow \left(\sum_{|\mathbf{u}| \in D_q} |\mathbf{u}|^2 \right)^{-1} \sum_{|\mathbf{u}| \in D_q} |\mathbf{u}|^2 C_f^j(\mathbf{u}) \quad (43)$$

$$P_{\text{HI}}(\mathbf{u}, \eta) \leftarrow \left(\sum_{\substack{|\mathbf{u}| \in D_q \\ \eta \in D_z}} \eta |\mathbf{u}|^2 \right)^{-1} \sum_{\substack{|\mathbf{u}| \in D_q \\ \eta \in D_z}} \eta |\mathbf{u}|^2 P_{\text{HI}}(\mathbf{u}, \eta) \quad (44)$$

where D_q is the set of \mathbf{u} values contributing to bin q and D_z is the set of η values contributing to bin z . Alternatively, if the 21-cm signal is expected to be highly isotropic, it is also appropriate to perform bin averages by averaging over spherical shells of constant $|k|$ to obtain the spherically averaged HI power spectrum. The bin-averaging schemes mentioned above are especially necessary at the beginning of iterations with a poor initial guess since bin averaging can smooth out any unreasonable values and highly suppress foreground contamination that is expected to significantly increase in strength towards low k_\parallel and would result in a substantial overestimate of HI power spectrum. One has to keep using the bin averaging process until the derived amplitude of the HI signal in each bin converges to a reasonable range.

In addition, although the bin-widths can be chosen arbitrarily, an appropriate choice can reduce the correlation between the band-power estimates while keeping accurate detection on the structure of the power spectrum. The minimum bin-width can be approximated by the typical correlation length in Fourier space. For an interferometer, one may choose a bin-width of perpendicular modes, $\Delta \mathbf{u}$, greater than the characteristic width of the Fourier transformed primary beam pattern $A(\mathbf{x})$, and a bin-width of parallel modes, $\Delta \eta$, roughly greater than the inverse of the bandwidth.

The EM algorithm iterates until there is no significant change in the likelihood. The stopping criterion of the iterative process is here set in terms of the relative change of the bin-averaged HI power spectrum in the last iterations, i.e., $|P_{\text{HI}}^{n+1} - P_{\text{HI}}^n| / |P_{\text{HI}}^n| < 10^{-3}$ for all bins, and a typical number of iterations is a few hundred.

2.5. The initial guess

Since the EM algorithm is a hill-climbing approach, the searching procedure may fail to reach a global maximum and instead converge on a local maximum if the likelihood functions are not convex. In this study, we adapt a commonly used strategy to solve this problem. We try many different initial values varied in reasonable ranges and choose the solution that has the highest converged likelihood value. After a careful investigation, we chose the following algorithm for initialization. For the initial value of \mathbf{M} , we perform PCA analysis de Oliveira-Costa et al. (2008) to obtain the n_c eigenvectors associated with the largest eigenvalues for the frequency-frequency covariance matrix estimated by averaging over all pixels of data, and use those vectors as the n_c columns of the mixing matrix \mathbf{M} . After the mixing matrix is initialized, the power spectra C_f^j are chosen to be the corresponding

diagonal elements in $\hat{\mathbf{x}}(\hat{\mathbf{x}})^\dagger$, which is solved using Eq. 23 and setting $\mathbf{C}^{-1} = 0$ such that there is no prior information about \mathbf{C} ; Eq. 23 reduces to a standard map-making equation. After initializing \mathbf{M} and C_f^j , we assume a flat power spectrum with amplitude comparable to the noise level as prior information about the HI signal. The simulations show that the HI signal reconstruction is quite insensitive to the initial guess for its power spectrum as long as it is not unreasonably large.

3. SIMULATIONS

We perform simulations to generate dirty sky map data cubes. These data cubes include realistic models for the cosmological 21-cm signal, several diffuse foreground components, and instrumental noise. As this is just the first test of HIEMICA, for simplicity we assume that the primary beam is unity for all frequency channels and assume complete uv-coverage. We leave detailed investigations about more realistic observations to future work.

Because an interferometer only measures temperature fluctuations around the mean and is insensitive to the mean value of brightness temperature, we set the mean of the 3D HI signal and the means of the foreground and noise maps at each frequency channel to zero.

The 3D dirty sky map is simulated in a box with 64^3 pixels (i.e., 64 pixels per side), covering a $30^\circ \times 30^\circ$ sky patch and spanning over 780 – 880 MHz at intervals of 1.59 MHz, corresponding to redshifts between 0.82 and 0.61. In comoving coordinates, this box corresponds to about $1341 \times 1341 \times 606 \text{ Mpc}^3$ and the size of each pixel is about $21 \times 21 \times 9.5 \text{ Mpc}^3$. The code was run using the best-fit cosmological parameters from the Planck measurements (Planck Collaboration et al. 2014b).

3.1. HI Signal

In cosmology the power spectrum is typically represented in the $(\mathbf{k}_\perp, k_\parallel)$ comoving coordinates. If the observed frequency band is small enough (i.e., probing a small range in redshifts) and one uses the flat-sky approximation, there is a linear mapping between these variables:

$$\mathbf{u} = \frac{\mathbf{k}_\perp D_c(z)}{2\pi}; \eta \approx \frac{c(1+z)^2}{2\pi H_0 \nu_{21} E(z)} k_\parallel, \quad (45)$$

where $E(z) \equiv \sqrt{\Omega_m(1+z)^3 + \Omega_\Lambda}$, ν_{21} is the rest frequency of the 21-cm line, D_c is the transverse comoving distance, z is the redshift of the observation, H_0 is the Hubble parameter, c is the speed of light, and Ω_m and Ω_Λ are the normalized matter and dark energy density, respectively. The angular wavenumber \mathbf{k}_\perp and the parallel wavenumber k_\parallel are the components of the wavenumber \mathbf{k} perpendicular and parallel to the line-of-sight direction, respectively. Therefore inserting Eq. 45 into 13, we obtain the relation of the power spectrum defined under the different coordinates:

$$P_{\text{HI}}(\mathbf{k}_\perp, k_\parallel) = \frac{c(1+z)^2}{H_0 \nu_{21} E(z)} P_{\text{HI}}(\mathbf{u}, \eta). \quad (46)$$

The 21-cm brightness temperature and the corresponding 3D power spectrum can be written as:

$$P_{\text{HI}}(k) = (\bar{T}_{\text{HI}}(z))^2 b^2(k, z) D^2(z) P_{\text{cdm}}(k) \quad (47)$$

$$\bar{T}_{21}(z) \simeq 0.084 \text{ mK} \frac{(1+z)^2}{E(z)} \frac{\Omega_B}{0.044} \frac{f_{\text{HI}}(z)}{0.01} \quad (48)$$

	$A(\text{mK}^2)$	β	α	ξ	$\Delta\alpha$
extragalactic point sources	57.0	1.1	2.07	1.0	0.2
extragalactic free-free	0.014	1.0	2.10	35	0.03
Galactic synchrotron	700	2.4	2.80	4.0	0.15
Galactic free-free	0.088	3.0	2.15	35	0.03

TABLE 1
FOREGROUND MODEL PARAMETERS FOR ANGULAR POWER SPECTRUM $C_\ell(\nu, \nu')$ USED IN EQ. 51.

where $b(k, z)$ is the bias parameter and P_{cdm} is the cold dark matter power spectrum at the present day. Ω_B is the baryon density fraction. $D(z)$ is the growth factor for dark matter perturbations defined such that $D(0) = 1$. For the purpose of this paper, we assume $b = 1$ over redshift and scale. For simplicity, we also neglect the effects of redshift-space distortions caused by the peculiar velocities of HI clouds and galaxies in the HI power spectrum since the ICA-based approach is insensitive to the detailed shape of the power spectrum. The HI mass fraction is expected to increase with redshift and we assume a linear dependence: $f_{\text{HI}}(z) = 0.008(1+z)$. The matter power spectrum P_{cdm} with the baryon acoustic oscillations (BAO) can be parametrized according to the simple empirical formula:

$$\frac{P_{\text{cdm}}(k_\perp, k_\parallel)}{P_{\text{ref}}(k_\perp, k_\parallel)} = 1 + \quad (49)$$

$$A k \exp(-(k/\tau)^\alpha) \sin\left(2\pi\sqrt{\frac{k_\perp^2}{k_{\text{BAO}\perp}^2} + \frac{k_\parallel^2}{k_{\text{BAO}\parallel}^2}}\right) \quad (50)$$

where $k = \sqrt{k_\perp^2 + k_\parallel^2}$, the parameters A , α and τ are adjusted according to the formula presented in Eisenstein & Hu (1998). $P_{\text{ref}}(k)$ is the smooth “no-wiggles” power spectrum at $z = 0$, which can be computed from the fitting formula given by Eisenstein & Hu (1998).

The parameters $k_{\text{BAO}\perp}$ and $k_{\text{BAO}\parallel}$ are the sinusoidal scales in the radial and transverse directions in k -space. We choose the following values for these parameters used in this paper: $A = 1.0$, $\tau = 0.1 \text{ hMpc}^{-1}$, $\alpha = 1.4$ and $k_{\text{BAO}\perp} = k_{\text{BAO}\parallel} = 0.060 \text{ hMpc}^{-1}$.

3.2. Foregrounds

We model the foregrounds as isotropic random Gaussian fields described by angular power spectra $C_\ell(\nu, \nu')$ based on Santos et al. (2005). In this paper, we assume that bright resolved point sources have been removed accurately and consider the dominant four diffuse components: Galactic synchrotron emission, Galactic and extra-galactic free-free emission, and extragalactic radio point sources. The angular power spectrum of each source takes the generic form:

$$C_f(\ell, \nu, \nu') = A \left(\frac{1000}{\ell}\right)^\beta \left(\frac{\nu_f^2}{\nu\nu'}\right)^{2\alpha} \exp\left(-\frac{\ln^2(\nu/\nu')}{2\xi^2}\right) \quad (51)$$

where ν_f is the reference frequency with $\nu_f = 130 \text{ MHz}$. We list the parameters of the foreground models used in this paper in Tab. 1. Based on such models, Fig. 1 shows the angular power spectrum of each foreground contribution for one realization. The corresponding maps are shown in Fig. 2.

Following Davies et al. (2006), the frequency index α of real-world foregrounds varies slightly across the sky. For example, in the case of synchrotron emission, the direction-dependent spectral index reflects variations of the relativistic electrons density and Galactic magnetic field. In this paper, for each foreground component, we assume the indices in different directions to be Gaussian distributed with a mean of α and an rms of $\Delta\alpha$ as shown in Tab. 1, consistent with model parameters chosen in Liu & Tegmark (2012).

3.3. Noise

For each frequency channel, we assume Gaussian noise in the uv -plane with zero mean and that the noise maps at two different frequencies are uncorrelated. We make noise realizations in the uv -plane for a given noise power spectrum for each frequency. The thermal noise of the measurement in units of brightness temperature is given by (McQuinn et al. 2006; Ansari et al. 2012a)

$$P_N(\mathbf{u}, \nu) = \langle |\tilde{n}(\mathbf{u}, \nu)|^2 \rangle = \left(\frac{\lambda^2}{A_e} \right)^2 \frac{T_{sys}^2}{\Delta\nu \Delta t n_b} \quad (52)$$

$$\simeq 1.9 \times 10^{-7} \text{ mK}^2 (1 + z_\nu)^4 \left(\frac{T_{sys}}{50 \text{ K}} \right)^2$$

$$\times \left(\frac{10 \text{ m}}{A_e} \right) \left(\frac{10}{n_b} \right) \left(\frac{1 \text{ MHz}}{\Delta\nu} \right) \left(\frac{30 \text{ days}}{\Delta t} \right)$$

where z_ν is the corresponding redshift of the frequency ν , T_{sys} is the system temperature, n_b is the total number of redundant baselines within that u_k pixel with integration time Δt for each baseline, $\Delta\nu$ is the width of the frequency channel, and A_e is the effective area of each individual antenna in the array.

In order to keep the results as general as possible, we do not consider a specific configuration of the array; the main goal in this paper is not to discuss instrumental effects in 21-cm power spectrum measurements. Instead, we consider two simple scenarios in our simulation to illustrate the effectiveness of the presented foreground-cleaning approach. One scenario assumes a currently achievable noise level of $\sqrt{P_N} \simeq 1 \times 10^{-3}$ mK for each pixel of the uv -plane, and the other one assumes a noise level of about 2.2×10^{-4} mK for future measurements. The former has an averaged signal-to-noise ratio over the whole data cube about 1 while the latter is about 5.

4. APPLICATIONS TO SIMULATED DATA

In this Section we show the results obtained by applying the HIEMICA algorithm to the simulated data. There are two main findings, namely the map reconstruction and the power spectrum recovery. As the fluctuations of the cosmological 21-cm signal are expected to be isotropic, we focus on the recovery of the averaged 3D power spectrum in spherical shells of constant $|k|$. However, redshift-space distortions will slightly break spherical symmetry, and we will investigate the consequences in a future work. The estimated power spectra are obtained by averaging over the results from 10 independent realizations of the simulated data cubes and the associated statistical errors are obtained from their dispersions. Each data cube is for a fixed sky patch that includes the same simulated foregrounds combined with independent realizations of the instrumental noise and the HI signal.

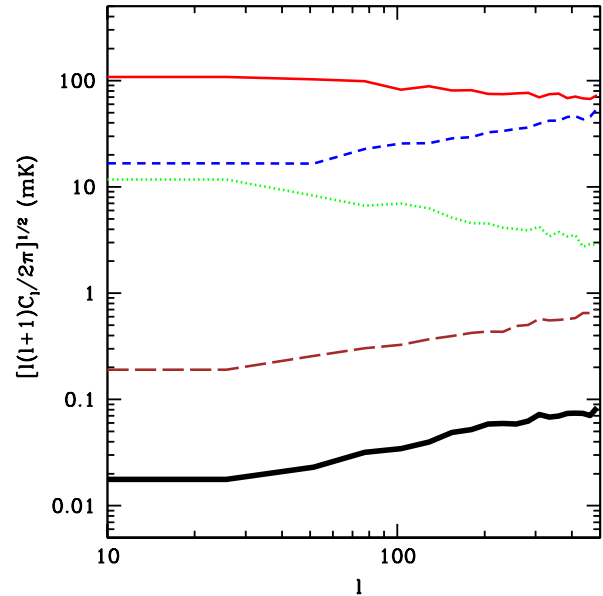


FIG. 1.— The angular power spectra derived from the simulated maps of the 21-cm signal (black-thick) and foregrounds at $z = 0.71$ ($\nu = 830$ MHz), with a bandwidth of $\Delta\nu = 1.59$ MHz. The foreground components include Galactic synchrotron radiation (red-solid), extragalactic point sources (blue-dashed), and Galactic free-free (green-dotted) and extragalactic free-free (brown long-dashed).

4.1. The Independent Components

Fig. 3 shows the four Wiener-filtered ICs at $\nu = 830$ MHz found by HIEMICA. The Wiener-filtered components are calculated by Eq. 23, and, the corresponding map of the i -th component at the frequency bin ν is obtained by multiplying such component by the (ν, i) -th entry of the mixing matrix, $M_{\nu, i}$. As seen from Figs. 2, 3 and 4, we find that only the first component is compatible with the input Galactic synchrotron contribution, while the others have no corresponding physical foregrounds and have unphysical spectral and spatial behaviours. These unphysical behaviours are due essentially to the fact that the realistic astrophysical components have significant cross-correlations while our independent component analysis assumes mutual independence between the source signals. Thus each separated component is just a mixture of all the source signals, but the sum of these ICs is essentially same as the total astrophysical signal, as seen in Figs. 5(a) and 5(b). Meanwhile, the Wiener-filtered maps of the 21-cm signal (bottom panel of Fig. 5) show the encouraging result that, the 21-cm signal, even in the case of $S/N=1$, is nearly perfectly recovered across the map by HIEMICA.

There is a slight difference between the low and high S/N cases (Figs. 5(d) and 5(e)) at small scales, but no apparent difference at large scales is found. The recovered amplitudes of fluctuations for $S/N=1$ are smaller than those for $S/N=5$ by several tens of percent, which is essentially due to noise suppression in Wiener-filtered maps that is $\propto S/(S+N)$, where S/N tends to be low at high- k modes. However, the large-scale fluctuations of the 21-cm signal in both cases are much larger than the noise amplitudes, resulting in almost identical large-scale fluctuations.

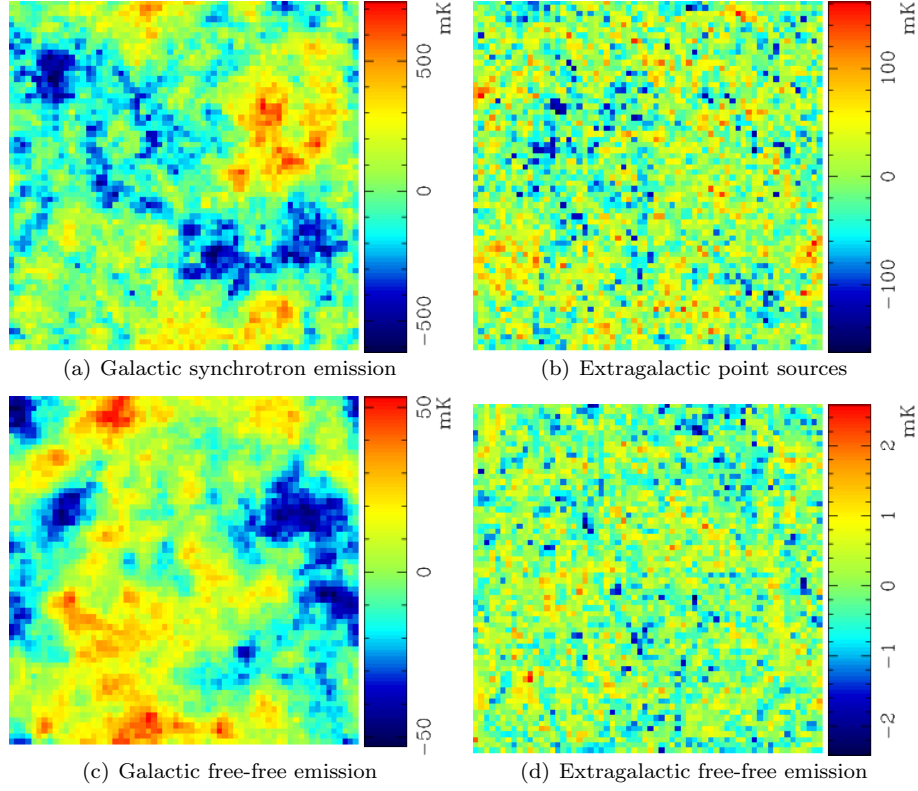


FIG. 2.— The simulated diffuse foreground components for a $30^\circ \times 30^\circ$ patch of sky at $\nu = 830$ MHz. We have subtracted the mean value (i.e. DC mode) to reflect an interferometric observation.

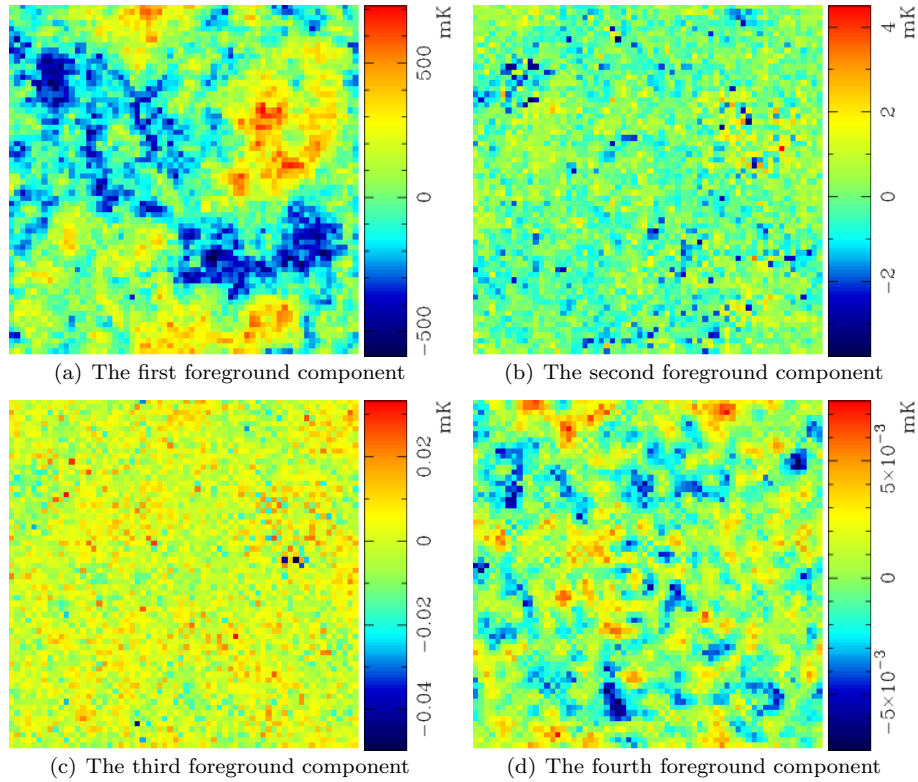


FIG. 3.— Wiener-filtered maps of independent components (ICs) reconstructed by applying HIEMICA to the simulated data cube with $S/N=5$ for the same sky patch and frequency as in Fig. 2. Notice that we do not expect the recovered maps to explicitly correspond to the input maps since these ICs are assumed to be mutually independent while the true physical components may have correlations.

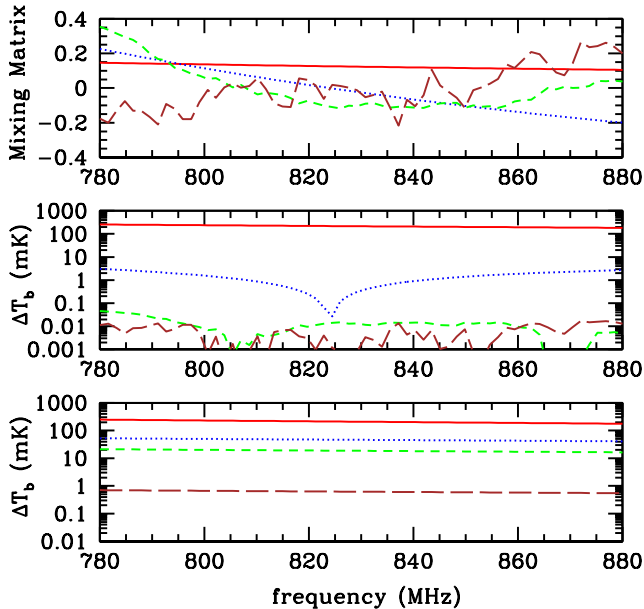


FIG. 4.— **Top panel:** HIEMICA-derived coefficients of each column of the mixing matrix, representing the frequency dependence of the ICs, where the red-solid line is for the first component, blue-dotted for the second, green-dashed for the third and brown long-dashed for the fourth. When applying HIEMICA to simulated data cubes, here we have assumed $n_c=4$ and $S/N=5$. **Middle panel:** the rms brightness temperatures of ICs calculated from the derived mixing matrix and their angular power spectra. **Bottom panel:** the rms brightness temperatures of the input physical foreground components including Galactic synchrotron radiation (red-solid), extragalactic point sources (blue-dotted), and Galactic free-free (green-dotted) and extragalactic free-free (brown long-dashed). As mentioned in Fig. 3, the temperature fluctuations of the ICs are not necessarily the same as the physical components.

4.2. Determination of the Number of ICs

When analyzing data with HIEMICA, we have to choose the number of Independent foreground Components (ICs) to be estimated. Our statistical framework for independent component analysis can provide a rigorous determination of this number by using an empirical statistical approach, the Likelihood Ratio Test (LRT). The LRT provides an objective criterion for selecting between possible models by using the ratio of the likelihoods on the data. Although in practice evaluating likelihood functions is a hard task, we can instead apply HIEMICA several times with successively increasing N_c until the recovered power spectra are essentially unchanged when more components are added into the analysis.

The LRT statistic is given by $LRT = 2 \ln \mathcal{L}_s - 2 \ln \mathcal{L}_c$, in which $\ln \mathcal{L}_s$ and $\ln \mathcal{L}_c$ denote the logs of the likelihood functions for a relatively simpler model s (so-called null hypothesis) with fewer parameters and a more complex model c (alternative hypothesis), respectively. Asymptotically, the LRT statistic follows a χ^2 distribution with degrees of freedom equal to the difference in the number of free parameters between the two models. In our case, in analysis of the data with 64 frequency channels and 30 ℓ -bins, adding an additional foreground component would increase the degrees of freedom by $64 + 30 = 94$.

The comparison of the fitting model between the models for $n_c = 1$ and 2 (i.e., labelled as $1 \rightarrow 2$ ICs in Tab. 2)

and $n_c = 2$ and 3 (as $2 \rightarrow 3$ ICs) shows that, from the values of LRT, $\ln L$ increases significantly as the number of ICs increased up to 3. The significance values ($p \simeq 0$) for S/N of 1 and 5 indicate that only one or two foreground components would result in an unsuccessful fit and the data strongly prefer three foreground components. Moreover, especially in the case of high S/N , there is no statistically significant difference when using more components, given the larger p values ($p > 0.01$). Here all the values of LRT and p are the median over 10 independent simulations. Therefore, it is recommended to use three components in the reconstructed foreground model for our specific simulation parameters in Tab. 1. It is worth noticing that, as mentioned by Alonso et al. (2015), the optimal number of ICs in fact strongly depends on the spectral smoothness of true foregrounds, characterized by a frequency correlation length ξ as defined in Eq. 51. In fact, a longer coherence length implies a smoother frequency spectrum for a physical foreground component. Consequently, a smaller value of ξ may require more ICs to successfully model such physical components and remove them accordingly.

4.3. 3D Power Spectrum Results

Fig. 6 shows the rapid convergence of the resulting HI power spectrum when successively increasing the number of ICs. It can be seen that the recovered power spectrum for IC number of 2 is slightly different from those for IC number of 3 and 4, but the 3 and 4 components result in almost the same power spectrum, which is consistent with the varying trend in the LRT. It can also be confirmed by comparing the reconstructed Wiener-filtered ICs in Fig 3. As the amplitude of fluctuations of the second component (see Fig. 3(b)) is indeed about 20 times larger than the input 21-cm signal (see Fig. 5(c)), the removal of the second component is therefore necessary. Furthermore, since the averaged amplitude of fluctuations of the third component (see Figs. 3(c) and 4) is about 10 times smaller than the 21-cm signal, the effects of such component in signal recovery are expected to be less important but can still lead to notable improvements at large scales as the corresponding 3D Fourier modes are not significantly smaller than those of the HI signal. Moreover, the amplitude of fluctuations of the fourth component (see Fig. 3(d)) is about two times smaller than the third one, leading to rather small effects in signal reconstruction.

The comparison between the estimated signal power spectra for $S/N=1$ and 5 in Fig. 6, we note that, for $k \lesssim 0.2 \text{ Mpc}^{-1}$ over which the averaged $S/N > 1$ for both cases, the impact of the noises at these scales tends to be negligibly small since the recovered power spectra follow the true spectrum very closely and there are no apparent differences between the two cases. On the other hand, the recovered power spectrum for $S/N=1$ begins gradually to overestimate the true one at $k > 0.2 \text{ Mpc}^{-1}$ by up to 12%, but for $S/N=5$ the recovered spectrum is almost same as the true one with only 2% deviation. The additional power in the former case is primarily because the large noise level causes a slight underestimate of the true foreground contributions. The “excess power” is barely significant; the simulation truth is still contained in the $1\text{-}\sigma$ error bars.

In Figs. 6(c) and 6(d), we zoom in on the region with

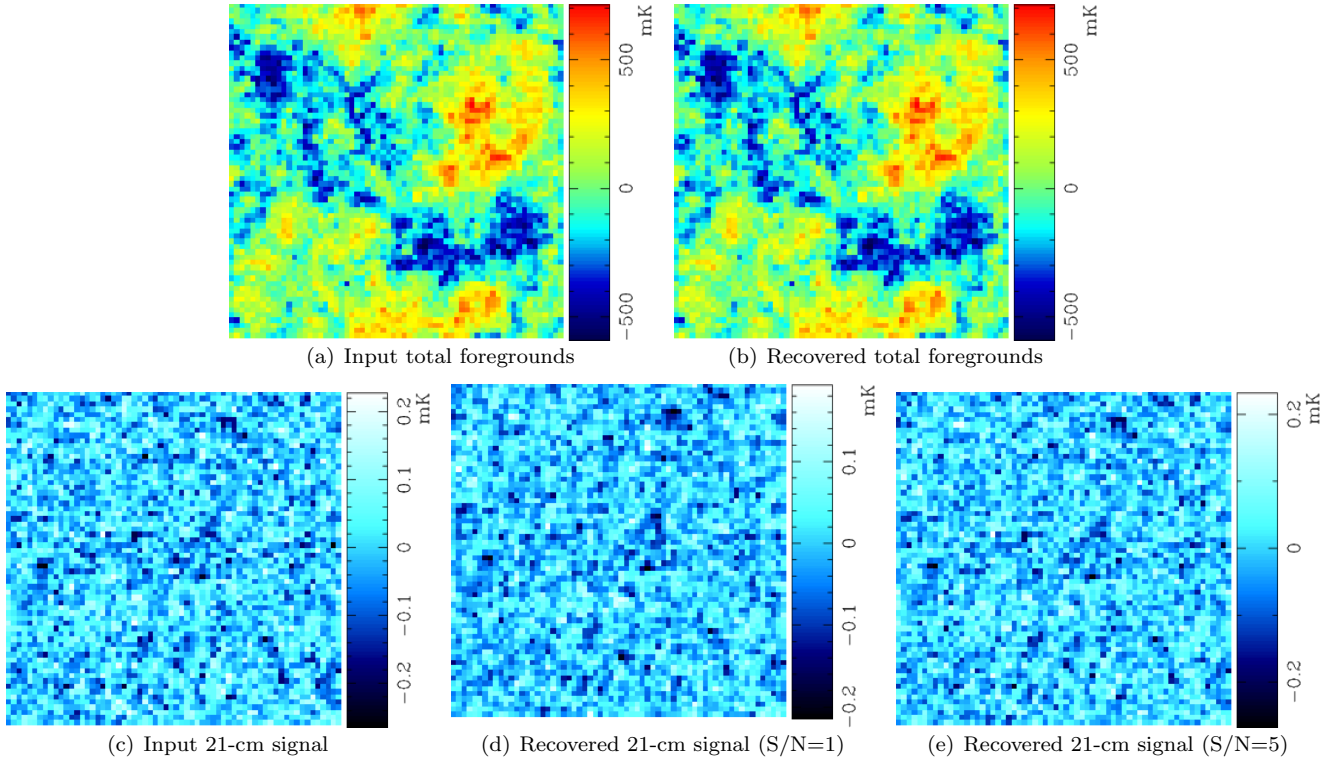


FIG. 5.— Same as Fig. 3, but for the total foregrounds and the 21-cm signal to clearly illustrate that a successful separation was achieved by the HIEMICA-cleaning process. There is an underestimate of the 21-cm signal in the case of $S/N=1$, which is caused by noise contamination as a Wiener-filtered map $\propto S/(S+N)$. An almost perfect recovery occurs over all scales for $S/N=5$.

BAO features to illustrate clearly the dependence of the power spectrum estimation on the number of ICs as well as on the noise amplitude. We can see that too few components, such as $n_c = 2$ in both cases of $S/N=1$ and 5, results in an overestimate on the power spectrum at scales of $0.1 \lesssim k \lesssim 0.2 \text{ Mpc}^{-1}$, since those two components can not completely describe the spectral and spatial properties of the foreground. As the number of ICs increases to 3 and 4, the resulting power spectra decrease and both rapidly converge to the true values. As the S/N drops down by a factor of 5, the measurement errors become larger since such errors can be approximately estimated by $\sigma(k) \propto \sqrt{P_{\text{HI}}(k) + P_N(k)}$. More importantly, one can see that, if $n_c = 3$ or 4, the recovered HI spectrum is well within the $2\text{-}\sigma$ error bars of the estimates.

Notice that we have applied a correction to convert the prediction intervals on the underlying power spectrum into error bars that approximate Bayesian power spectrum inference in the limit of low noise and small numbers of modes. The resulting upper and lower error bars are asymmetric. The detailed analytical estimations of such correction are presented in the Appendix. As a consequence, compared with the uncorrected errors, the upper error bars increased by about 60%, 40% and 20% respectively in the lowest three k -bins and the associated lower error bars decreased by about 40%, 15%, 10%. The differences are less than a few percent in the higher k -bins due to the large number of modes contained within those bins.

We can also see that the recovered cosmological signal decreases approximately monotonically with increasing number of ICs, especially for the low- k modes ($k \lesssim 0.1 \text{ Mpc}^{-1}$). This result may be interpreted as follows: (1)

since each recovered foreground component is a mixture of all the source signals, including the cosmological signal, a small amount of cosmological signal can leak into the reconstructed ICs, strongly biasing and reducing the amplitude of the derived cosmological signal; (2) the angular power spectrum of the dominant foreground component, Galactic synchrotron emission, has the largest amplitude for small ℓ (see Fig. 1), resulting in a strong contamination in the cosmological signal at large scales in the transverse direction while increasing the risk of leakage; (3) the performance of the ICs-based reconstruction also depends on the number of measured Fourier modes in each k -bin, and, in the simulated data the lowest and the second lowest k -bins contain about 500 and 50 times fewer modes than those in the high k -bins, respectively.

As discussed before, using the LRT, the three-component foreground model used in HIEMICA gives rise to the successful recovery of the HI power spectrum. In order to evaluate the recovery performance, we present the derived power spectrum in Fig. 7 in comparison with that from the existing commonly-used PCA technique. We can see that, after applying PCA to subtract the three dominant foreground components from the data cube, the residual apparently underestimates the HI power spectrum at large scales, which seems to be an over subtraction of the foregrounds. Compared with the input ones, the amplitudes in the first two k -bins are smaller by factors of 14.1 and 3.7 if $S/N=1$, and by factors of 16.1 and 4.1 if $S/N=5$, respectively, whereas the ratios of HIEMICA-derived power spectra and the input ones in those two bins are only about 91–92% if $S/N=1$ and 91–94% if $S/N=5$, respectively. In addition, the recovered amplitudes in those two k -bins are consistent with the input values within $2\text{-}\sigma$. The HIEMICA method

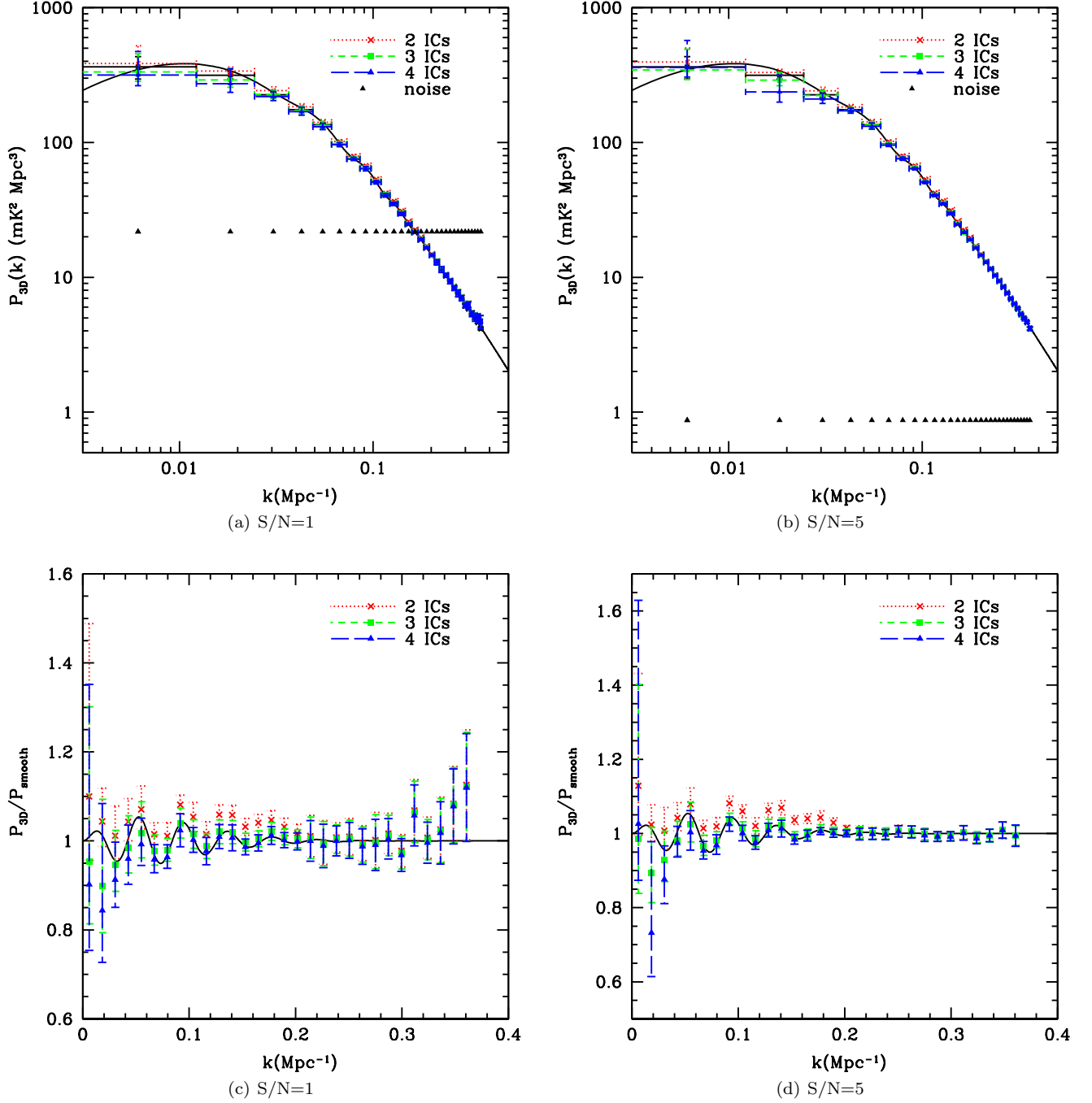


FIG. 6.— Dependence of the 21-cm signal recovery on the number of independent components, for the data with $S/N=1$ (left) and 5 (right). **Upper panels:** We show the spherically averaged three-dimensional power spectra of the simulated 21-cm signal (black), noise (black-dotted), reconstructed 21-cm signal for the HIEMICA algorithm run with the assumption of $n_c = 2$ (red-dotted), 3 (green-dashed) and 4 (blue long-dashed), respectively. Vertical bars indicate the $1-\sigma$ errors estimated from 10 realizations and horizontal bars the bin-width of $\Delta k = 0.0129$. The statistical uncertainties are typically smaller than the symbol sizes and mostly invisible. **Lower panels:** Same as upper panels, but showing the HI power spectrum divided by a smoothed spectrum without baryon acoustic oscillations to single out those oscillations.

TABLE 2
COMPARISON OF MODELS BY THE LIKELIHOOD RATIO TEST (LRT) FOR THE NUMBER OF INDEPENDENT FOREGROUND COMPONENTS.

Model	Δdof	LRT (S/N=1)	LRT (S/N=5)	p -value (S/N=1)	p -value (S/N=5)
1 \rightarrow 2 ICs	94	17558	20872	0 [†]	0 [†]
2 \rightarrow 3 ICs	94	196	280	$< 10^{-6\ddagger}$	$< 10^{-6\ddagger}$
3 \rightarrow 4 ICs	94	120	125	0.036	0.018
4 \rightarrow 5 ICs	94	96	108	0.42	0.15

[†] $p < 0.01$, typically used to assess the significance of LRT.

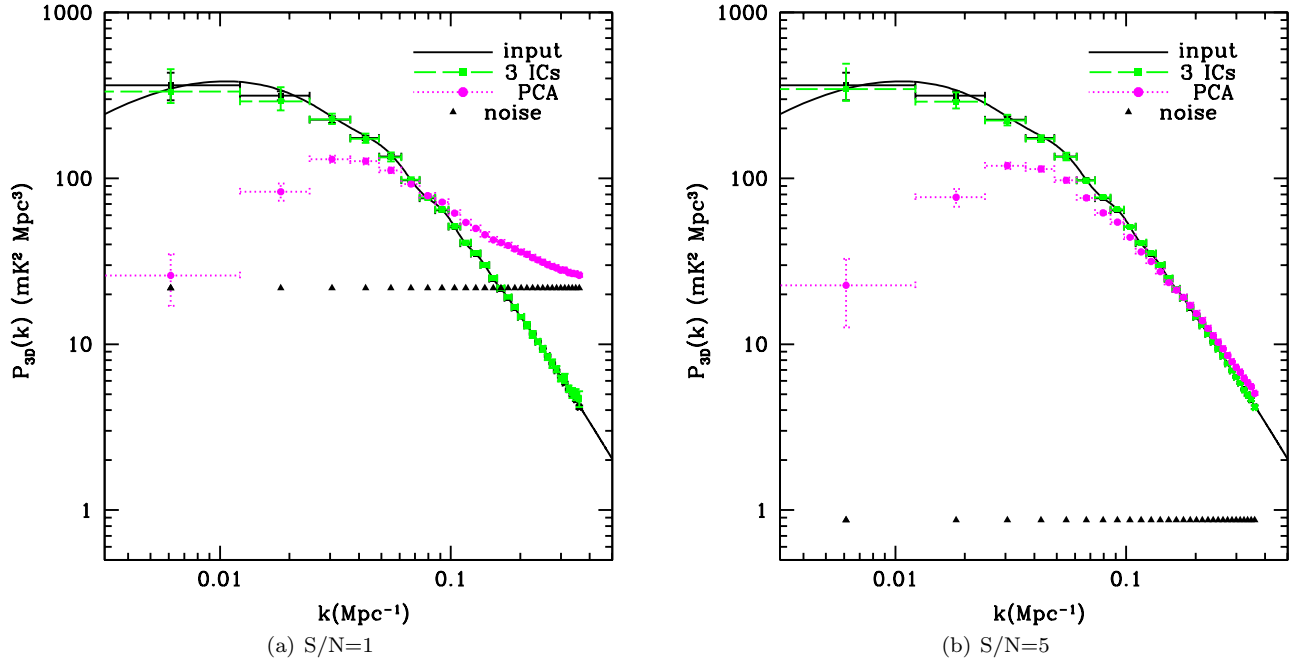


FIG. 7.— Same as Fig. 6, but for comparison of 21-cm power spectrum recovery based on the PCA technique (magenta-dotted) and the HIEMICA approach (green long-dashed), for the data with S/N=1 (left) and S/N=5 (right). The spherically averaged three-dimensional power spectra of the simulated 21-cm signal (black-solid), noise (black-dotted), HIEMICA-derived (assumed $n_c=3$) 21-cm signal and PCA-derived residuals from projecting out the first three dominant eigen-components based on the frequency-frequency covariance of data.

thus does a much better job at foreground removal and cosmological signal recovery at large scales.

More promisingly, we find that the HIEMICA approach provides an unbiased estimate of HI power spectrum for the high- k modes with $k \gtrsim 0.1 \text{ Mpc}^{-1}$ where S/N tends to be low, whereas the power spectrum of the PCA-derived residual in the noise-dominant regime seems to deviate strongly from the true signal, resulting in an overestimate of HI measurements.

Since the isotropic HI signal has the same statistical property as that of the assumed instrumental noise, commonly-used ICA-based approaches have trouble breaking the degeneracy, leaving residuals consisting of the reconstructed 21-cm signal, noise and fitting errors. As proposed in Chapman et al. (2012); Bonaldi & Brown (2015), after foreground cleaning by CCA or FASTICA, one has to manually subtract the noise power spectrum from that of the residual data cube and obtain a comparable amplitude with the true signal. However, those previous studies show that such a scheme can cause an overfitting problem and lead to some negative or zero-closed amplitudes of the residual power spectrum in noise-dominated k -modes, implying that a leakage of HI signal into the subtracted foreground components is likely to be serious in regions with low S/N. Some other schemes such as weighting maps with inverse noise variance to build the frequency covariance used in PCA (Alonso et al. 2015) could mitigate such noise contamination.

In contrast to the above-mentioned standard PCA- and ICA-based approaches, the proposed HIEMICA method in fact provides an unbiased and efficient estimate of Gaussian components, derived from the maximum likelihood principle. If the noise covariance matrix \mathbf{N} is exactly known, this method can optimally estimate the underlying 21-cm signal and foreground covariance matrices. It avoids any artificial effects that can occur when manually subtracting the noise power spectrum from the residuals after the foreground cleaning processes. Our simulations do confirm this, showing that the ensemble averaged recovered HI power spectrum converges to the true value even in low S/N regions where the noise is about an order of magnitude higher.

We note that our semi-blind approach requires exact knowledge of the noise in order to infer the 21-cm power spectrum accurately; an incorrect noise covariance matrix can bias the signal estimate. However, a relatively lower accuracy in modelling the noise covariance matrix may be sufficient for the power spectrum estimation if the uncertainty in the noise level is not greater than the signal. In general, the corresponding impacts on the estimated parameters can be evaluated by a Monte-Carlo approach. But qualitatively, the level of the uncertainty in the estimated 21-cm power spectrum is about of the same order of magnitude as that of the noise. A simple way to understand this is to consider that in the absence of instrumental effects, if the noise is assumed to be uncorrelated, the value of P_{HI} which maximizes the likelihood function can be approximately estimated by $P_{\text{HI}} \approx P_{\text{D}} - P_{\text{F}} - P_{\text{N}}$ where P_{D} , P_{F} and P_{N} are the power spectra of the data points, the foreground and noise, respectively. As the total amplitude of the foreground fluctuations is significantly larger than that of the noise fluctuations, one can expect that the estimate

of the foreground power spectrum roughly remains unchanged when slightly varying the input noise amplitude. Thus assuming an input noise with an amplitude either greater or smaller than the true one would lead to either an underestimate or an overestimate of the 21-cm power spectrum, i.e., $\Delta P_{\text{HI}} \approx -\Delta P_{\text{N}}$. In order to check the above estimate, we test the effect of using an input noise power spectrum with 10 – 50% larger or smaller amplitudes than the true one and find that the resulting changes ΔP_{HI} are indeed consistent with our expected variations ($-\Delta P_{\text{N}}$) within a factor of 2 over all scales. In addition, incorrect noise estimates would also bias the estimate for the PCA method, since the debiasing Monte Carlo simulations would give the wrong results if the noise level were misspecified.

In more realistic cases, an experiment may exhibit a low-level of cross-correlated Gaussian or non-Gaussian noise, both of which are unknown. In the former case, similar to the above qualitative estimation, we expect that the small level of correlation will slightly bias the signal estimate but not significantly affect it. This is because, as mentioned above, (1) the HI power spectrum is approximately determined by the difference among the diagonal components of the data point, the foreground, and noise model, not by their off-diagonal ones, and (2) the estimate of the foreground is unchanged or insensitive to the noise model. In the latter case, as the variance of the non-Gaussian noise is non-negative, the HI power spectrum is thus likely to be slightly overestimated. We will investigate the impact of noise misspecification in a future work.

Also, this blind component separation algorithm makes no assumptions about the spectral information of the foreground and can naturally include all instrumental effects so as to be able to coherently estimate all source components and the associated errors. More detailed comparisons between HIEMICA and other approaches for data in the presence of realistic instrumental effects will be made in future work.

5. CONCLUSIONS

In this paper we present a non-parametric source separation algorithm, HIEMICA, which is an extension to the 3D version of the SMICA method proposed by Snoussi et al. (2002); Delabrouille et al. (2003) for source identification in noisy mixtures of CMB maps. The HIEMICA algorithm is a fully Bayesian framework to infer the 3D power spectrum and maps of the underlying HI signal and the spatial power spectra and the frequency dependence (i.e. mixing matrix) of uncorrelated foreground components. We adapt the EM algorithm to efficiently maximize the likelihood function for unknown parameter estimates. As the statistical properties of the cosmological signal are significantly different from those of each astrophysical source, the spectral matching method in principle can blindly separate the HI signal from highly foreground-contaminated maps.

The simulations show that HIEMICA is able to successfully reconstruct the HI 3D power spectrum across all scales and is much more robust than the PCA approach, which removes some of the 21-cm signal at large scales ($k \lesssim 0.1 \text{ Mpc}^{-1}$) and overestimates the HI power spectrum in the noise-dominated regions where noise leaks into the estimated signal. To evaluate the impact of

the number of ICs in signal recovery we adapt the LRT method to rigorously assess the likelihood fit so as to determine the optimal number of foreground components used in the HIEMICA analysis. We find for the number of ICs higher than 3, a number comparable to that of the simulated astrophysical components, the reconstruction does not gain significant improvements and converges towards a stable result consistent with the input HI power spectrum.

Although our results are quite promising, the simulated observations in this study are idealized, with no instrumental effects. As instrument effects such as the frequency-dependent uv sampling and primary beam can significantly complicate the foreground removal and bias the signal estimate, it is important to test HIEMICA in

more detail in the context of specific experiment set-ups in future work.

ACKNOWLEDGMENTS

We acknowledge use of the open-source PETSc library (Balay et al. 2014a,b, 1997) and FFTW (Frigo & Johnson 2005). LZ would like to thank Flavien Vansyngel for useful comments on early drafts. LZ and PT acknowledge support from NSF awards IIS 1250720 (‘Big Data’ program) and AST 1216525. PMS is supported by the INFN IS PD51 “Indark”. We would like to thank an anonymous referee for valuable suggestions which helped us to significantly improve this paper.

REFERENCES

- Alonso, D., Bull, P., Ferreira, P. G., & Santos, M. G. 2015, *MNRAS*, 447, 400
- Anderson, L., Aubourg, E., Bailey, S., Bizyaev, D., Blanton, M., et al. 2013, *Mon.Not.Roy.Astron.Soc.*, 427, 3435
- Ansari, R., Campagne, J. E., Colom, P., Le Goff, J. M., Magneville, C., Martin, J. M., Moniez, M., Rich, J., & Yèche, C. 2012a, *A&A*, 540, A129
- Ansari, R., Campagne, J.-E., Colom, P., Magneville, C., Martin, J.-M., Moniez, M., Rich, J., & Yèche, C. 2012b, *Comptes Rendus Physique*, 13, 46
- Balay, S., Abhyankar, S., Adams, M. F., Brown, J., Brune, P., Buschelman, K., Eijkhout, V., Gropp, W. D., Kaushik, D., Knepley, M. G., McInnes, L. C., Rupp, K., Smith, B. F., & Zhang, H. 2014a, *PETSc Users Manual*, Tech. Rep. ANL-95/11 - Revision 3.5, Argonne National Laboratory
- . 2014b, *PETSc Web page*, <http://www.mcs.anl.gov/petsc>
- Balay, S., Gropp, W. D., McInnes, L. C., & Smith, B. F. 1997, in *Modern Software Tools in Scientific Computing*, ed. E. Arge, A. M. Bruaset, & H. P. Langtangen (Birkhäuser Press), 163–202
- Bonaldi, A. & Brown, M. L. 2015, *MNRAS*, 447, 1973
- Bowman, J. D., Morales, M. F., & Hewitt, J. N. 2006, *ApJ*, 638, 20
- Bowman, J. D., Morales, M. F., & Hewitt, J. N. 2009, *Astrophys.J.*, 695, 183
- Bull, P., Ferreira, P. G., Patel, P., & Santos, M. G. 2015, *ApJ*, 803, 21
- Cardoso, J.-F., Martin, M., Delabrouille, J., Betoule, M., & Patanchon, G. 2008, *IEEE Journal of Selected Topics in Signal Processing*, 2, 735
- Chang, T., Pen, U., Bandura, K., & Peterson, J. B. 2010, *Nature*, 466, 463
- Chang, T.-C., Pen, U.-L., Peterson, J. B., & McDonald, P. 2008, *Physical Review Letters*, 100, 091303
- Chapman, E., Abdalla, F. B., Harker, G., Jelić, V., Labropoulos, P., Zaroubi, S., Brentjens, M. A., de Bruyn, A. G., & Koopmans, L. V. E. 2012, *MNRAS*, 423, 2518
- Chen, X. 2012, *International Journal of Modern Physics Conference Series*, 12, 256
- Chen, X.-L. & Miralda-Escude, J. 2004, *Astrophys.J.*, 602, 1
- Colless, M., Peterson, B. A., Jackson, C., Peacock, J. A., Cole, S., Norberg, P., Baldry, I. K., Baugh, C. M., Bland-Hawthorn, J., Bridges, T., Cannon, R., Collins, C., Couch, W., Cross, N., Dalton, G., De Propris, R., Driver, S. P., Efstathiou, G., Ellis, R. S., Frenk, C. S., Glazebrook, K., Lahav, O., Lewis, I., Lumsden, S., Maddox, S., Madgwick, D., Sutherland, W., & Taylor, K. 2003, *ArXiv Astrophysics e-prints*
- Cooray, A. & Furlanetto, S. R. 2004, *ApJ*, 606, L5
- Datta, A., Bowman, J. D., & Carilli, C. L. 2010, *ApJ*, 724, 526
- Davies, R. D., Dickinson, C., Banday, A. J., Jaffe, T. R., Górski, K. M., & Davis, R. J. 2006, *MNRAS*, 370, 1125
- de Oliveira-Costa, A., Tegmark, M., Gaensler, B. M., Jonas, J., Landecker, T. L., & Reich, P. 2008, *MNRAS*, 388, 247
- Delabrouille, J., Cardoso, J.-F., & Patanchon, G. 2003, *MNRAS*, 346, 1089
- Di Matteo, T., Ciardi, B., & Miniati, F. 2004, *MNRAS*, 355, 1053
- Di Matteo, T., Perna, R., Abel, T., & Rees, M. J. 2002, *ApJ*, 564, 576
- Dillon, J. S., Liu, A., & Tegmark, M. 2013, *Phys. Rev. D*, 87, 043005
- Drinkwater, M. J., Jurek, R. J., Blake, C., Woods, D., Pimblett, K. A., et al. 2010, *Mon.Not.Roy.Astron.Soc.*, 401, 1429
- Eisenstein, D. J. & Hu, W. 1998, *ApJ*, 496, 605
- Fessler, J. A. & Sutton, B. P. 2003, *IEEE Trans. Signal Process.*, 51, 560
- Frigo, M. & Johnson, S. G. 2005, *Proceedings of the IEEE*, 93, 216, special issue on “Program Generation, Optimization, and Platform Adaptation”
- Furlanetto, S., Oh, S. P., & Briggs, F. 2006, *Phys.Rept.*, 433, 181
- Gleser, L., Nusser, A., & Benson, A. J. 2008, *MNRAS*, 391, 383
- Harker, G., Zaroubi, S., Bernardi, G., Brentjens, M. A., de Bruyn, A. G., Ciardi, B., Jelić, V., Koopmans, L. V. E., Labropoulos, P., Mellema, G., Offringa, A., Pandey, V. N., Pawlik, A. H., Schaye, J., Thomas, R. M., & Yatawatta, S. 2010, *MNRAS*, 405, 2492
- Harker, G., Zaroubi, S., Bernardi, G., Brentjens, M. A., de Bruyn, A. G., Ciardi, B., Jelić, V., Koopmans, L. V. E., Labropoulos, P., Mellema, G., Offringa, A., Pandey, V. N., Schaye, J., Thomas, R. M., & Yatawatta, S. 2009a, *MNRAS*, 397, 1138
- Harker, G., Zaroubi, S., Thomas, R. M., Jelić, V., Labropoulos, P., Mellema, G., Iliev, I. T., Bernardi, G., Brentjens, M. A., de Bruyn, A. G., Ciardi, B., Koopmans, L. V. E., Pandey, V. N., Pawlik, A. H., Schaye, J., & Yatawatta, S. 2009b, *MNRAS*, 393, 1449
- Hinshaw, G., Larson, D., Komatsu, E., et al. 2013, *ApJS*, 208, 19
- Hyvärinen, A. 1999, *IEEE Trans. On Neural Networks*, 10, 626
- Hyvärinen, A. & Oja, E. 2000, *Neural Netw.*, 13, 411
- Jeffreys, H. 1961, *Theory of Probability* (3rd Edition) (New York: Oxford University Press)
- Jelić, V., Zaroubi, S., Labropoulos, P., Thomas, R. M., Bernardi, G., Brentjens, M. A., de Bruyn, A. G., Ciardi, B., Harker, G., Koopmans, L. V. E., Pandey, V. N., Schaye, J., & Yatawatta, S. 2008, *MNRAS*, 389, 1319
- Jewell, J., Levin, S., & Anderson, C. H. 2004, *ApJ*, 609, 1
- Kamionkowski, M., Kosowsky, A., & Stebbins, A. 1997, *Physical Review Letters*, 78, 2058
- Kinney, W. H., Kolb, E. W., Melchiorri, A., & Riotto, A. 2006, *Phys. Rev. D*, 74, 023502
- Liu, A., Parsons, A. R., & Trott, C. M. 2014a, *Phys. Rev. D*, 90, 023018
- . 2014b, *Phys. Rev. D*, 90, 023019
- Liu, A. & Tegmark, M. 2012, *MNRAS*, 419, 3491
- Liu, A., Tegmark, M., Bowman, J., Hewitt, J., & Zaldarriaga, M. 2009a, *MNRAS*, 398, 401
- Liu, A., Tegmark, M., & Zaldarriaga, M. 2009b, *MNRAS*, 394, 1575
- Loeb, A. & Wyithe, J. S. B. 2008, *Physical Review Letters*, 100, 161301
- Loeb, A. & Zaldarriaga, M. 2004, *Physical Review Letters*, 92, 211301
- Madau, P., Meiksin, A., & Rees, M. J. 1997, *Astrophys.J.*, 475, 429
- Maldacena, J. M. 2003, *JHEP*, 0305, 013
- Mao, Y., Tegmark, M., McQuinn, M., Zaldarriaga, M., & Zahn, O. 2008, *Phys.Rev.*, D78, 023529

- Masui, K. W., Switzer, E. R., Banavar, N., Bandura, K., Blake, C., Calin, L.-M., Chang, T.-C., Chen, X., Li, Y.-C., Liao, Y.-W., Natarajan, A., Pen, U.-L., Peterson, J. B., Shaw, J. R., & Voytek, T. C. 2013, *ApJ*, 763, L20
- McQuinn, M., Zahn, O., Zaldarriaga, M., Hernquist, L., & Furlanetto, S. R. 2006, *ApJ*, 653, 815
- Morales, M. F., Bowman, J. D., & Hewitt, J. N. 2006, *ApJ*, 648, 767
- Morales, M. F., Hazelton, B., Sullivan, I., & Beardsley, A. 2012, *ApJ*, 752, 137
- Morales, M. F. & Hewitt, J. 2004, *ApJ*, 615, 7
- Morales, M. F. & Wyithe, J. S. B. 2010, *Annual Review of Astronomy and Astrophysics*, 48, 127
- Nan, R., Li, D., Jin, C., Wang, Q., Zhu, L., Zhu, W., Zhang, H., Yue, Y., & Qian, L. 2011, *International Journal of Modern Physics D*, 20, 989
- Oh, S. P. & Mack, K. J. 2003, *MNRAS*, 346, 871
- Parsons, A. R., Pober, J. C., Aguirre, J. E., Carilli, C. L., Jacobs, D. C., & Moore, D. F. 2012, *ApJ*, 756, 165
- Peiris, H. V., Komatsu, E., Verde, L., Spergel, D. N., Bennett, C. L., Halpern, M., Hinshaw, G., Jarosik, N., Kogut, A., Limon, M., Meyer, S. S., Page, L., Tucker, G. S., Wollack, E., & Wright, E. L. 2003, *ApJS*, 148, 213
- Pen, U.-L., Staveley-Smith, L., Peterson, J. B., & Chang, T.-C. 2009, *MNRAS*, 394, L6
- Peterson, J. B., Bandura, K., & Pen, U. L. 2006, in *Proceedings, 41st Recontres de Moriond, 2006 Contents and structure of the universe*, 283–289
- Petrovic, N. & Oh, S. P. 2011, *MNRAS*, 413, 2103
- Planck Collaboration, Ade, P. A. R., Aghanim, N., Armitage-Caplan, C., Arnaud, M., Ashdown, M., Atrio-Barandela, F., Aumont, J., Baccigalupi, C., Banday, A. J., & et al. 2014a, *A&A*, 571, A12
- Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2014b, *A&A*, 571, A16
- Pober, J. C., Liu, A., Dillon, J. S., Aguirre, J. E., Bowman, J. D., Bradley, R. F., Carilli, C. L., DeBoer, D. R., Hewitt, J. N., Jacobs, D. C., McQuinn, M., Morales, M. F., Parsons, A. R., Tegmark, M., & Werthimer, D. J. 2014, *ApJ*, 782, 66
- Pober, J. C., Parsons, A. R., Aguirre, J. E., Ali, Z., Bradley, R. F., Carilli, C. L., DeBoer, D., Dexter, M., Gugliucci, N. E., Jacobs, D. C., Klima, P. J., MacMahon, D., Manley, J., Moore, D. F., Stefan, I. I., & Walbrugh, W. P. 2013a, *ApJ*, 768, L36
- Pober, J. C., Parsons, A. R., DeBoer, D. R., McDonald, P., McQuinn, M., Aguirre, J. E., Ali, Z., Bradley, R. F., Chang, T.-C., & Morales, M. F. 2013b, *AJ*, 145, 65
- Pritchard, J. R. & Loeb, A. 2012, *Reports on Progress in Physics*, 75, 086901
- Santos, M. G., Cooray, A., & Knox, L. 2005, *ApJ*, 625, 575
- Seo, H.-J. & Eisenstein, D. J. 2003, *ApJ*, 598, 720
- Shaw, J. R., Sigurdson, K., Pen, U.-L., Stebbins, A., & Sitwell, M. 2014, *ApJ*, 781, 57
- Smoot, G. F., Bennett, C. L., Kogut, A., et al. 1992, *ApJ*, 396, L1
- Snoussi, H., Patanchon, G., Macías-Pérez, J. F., Mohammad-Djafari, A., & Delabrouille, J. 2002, in *American Institute of Physics Conference Series*, Vol. 617, *Bayesian Inference and Maximum Entropy Methods in Science and Engineering*, ed. R. L. Fry, 125–140
- Sutter, P. M., Wandelt, B. D., & Malu, S. S. 2012, *ApJS*, 202, 9
- Switzer, E. R., Masui, K. W., Bandura, K., Calin, L.-M., Chang, T.-C., Chen, X.-L., Li, Y.-C., Liao, Y.-W., Natarajan, A., Pen, U.-L., Peterson, J. B., Shaw, J. R., & Voytek, T. C. 2013, *MNRAS*, 434, L46
- Trott, C. M., Wayth, R. B., & Tingay, S. J. 2012, *ApJ*, 757, 101
- Vansyngel, F., Wandelt, B. D., Cardoso, J.-F., & Benabed, K. 2014, *ArXiv e-prints*
- Vedantham, H., Udaya Shankar, N., & Subrahmanyam, R. 2012, *ApJ*, 745, 176
- Wandelt, B. D., Larson, D. L., & Lakshminarayanan, A. 2004, *Phys. Rev. D*, 70, 083511
- Wang, X., Chen, X., Zheng, Z., Wu, F., Zhang, P., & Zhao, Y. 2009, *MNRAS*, 394, 1775
- Wang, X., Tegmark, M., Santos, M. G., & Knox, L. 2006, *ApJ*, 650, 529
- Wolz, L., Abdalla, F. B., Blake, C., Shaw, J. R., Chapman, E., & Rawlings, S. 2014, *MNRAS*, 441, 3271
- Wyithe, S., Loeb, A., & Geil, P. 2008, *MNRAS*, 383, 1195
- Xu, Y., Wang, X., & Chen, X. 2015, *ApJ*, 798, 40
- York, D. G., Adelman, J., Anderson, Jr., J. E., et al. 2000, *AJ*, 120, 1579
- Zaldarriaga, M., Furlanetto, S. R., & Hernquist, L. 2004, *ApJ*, 608, 622
- Zaldarriaga, M. & Seljak, U. 1998, *Phys. Rev. D*, 58, 023003

APPENDIX

We here present a derivation of an approximate expression for the posterior distribution of the signal power spectrum in terms of the sampling distribution of the estimator so as to calculate Bayesian credible intervals of the inferred parameter. The sampling distribution of the estimator in our case can be derived from a Monte Carlo (MC) distribution of the Maximum Likelihood Estimator (MLE) of the signal power spectrum P .

The power spectrum is simply the variance of the Fourier mode coefficients. For simplicity, if we assume that the all the Fourier modes are independent and the variance of all the Fourier mode coefficients within the bin is the same, the posterior distribution of P given data d_i for $i = 1, \dots, n$ can be expressed in an ideal noise-free case as

$$p(P|d) \propto \mathcal{L}(d|P)p(P) \propto P^{-\frac{n}{2}} e^{-\frac{1}{2} \sum_{i=1}^n d_i^2 / P} P^{-1}, \quad (1)$$

where the non-informative Jeffreys prior (Jeffreys 1961), $p(P) = P^{-1}$, has been applied on the power spectrum. As we know, the MLE $\hat{P} = \frac{1}{n} \sum_{i=1}^n d_i^2$ is unbiased over data sets drawn from

$$\langle \hat{P} \rangle = P \quad (2)$$

and has variance

$$\langle (\Delta \hat{P})^2 \rangle = \frac{2}{n} P^2. \quad (3)$$

If we want to construct an approximate Bayesian posterior from MC samples from the sampling density of the MLE \hat{P} and the effective number of degrees of freedom is unknown, then we can determine it from the mean and variance of \hat{P} using

$$\tilde{n} = 2 \frac{\langle \hat{P} \rangle}{\langle (\Delta \hat{P})^2 \rangle}. \quad (4)$$

Assuming that noise is subdominant, inserting Eq. 4 into Eq. 1, we can use \hat{P} and the effective number of degrees \tilde{n} of freedom to approximate the posterior for P as

$$p(P|d) \propto e^{-\frac{1}{2} \tilde{n} \hat{P} / P} / P^{\frac{\tilde{n}+2}{2}}, \quad (5)$$

from which the asymmetric credible interval can be derived.